Prediction of Long Term Living Donor Kidney Graft Outcome: Comparison Between Rule Based, Decision Tree and Linear Regression

Article *in* International Journal of Advanced Research in Computer Science · April 2015

CITATIONS
7

READS
479

4 authors, including:

Ahmed I Akl
Urology and Nephrology Center
111 PUBLICATIONS **694** CITATIONS

SEE PROFILE

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)
Vol. 3, Issue 2 (Apr. - Jun. 2015)
International Journal of Advanced Research in Computer Science & Technology (IJARCST 2015)

Prediction of Long Term Living Donor Kidney Graft Outcome: Comparison Between Rule Based, Decision Tree and Linear Regression

[I]Maha Fouad, [II]Dr. Mahmoud M. Abd Ellatif, [III]Prof.Mohamed Hagag , [IV]Dr.Ahmed Akl
[I]Dept. of Info. Systems, Faculty of Computer and Information Sc., Mansoura Uni., Mansoura, Egypt
[II,III]Faculty of Computer and Information Sciences, Helwan University, Egypt
[IV]Urology & Nephrology Centre, Mansoura University, Mansoura, Egypt

Abstract
Predicting the outcome of a graft transplant with high level of accuracy is a challenging task In medical fields and Data Mining has a great role to answer the challenge. The goal of this study is to compare the performances and features of data mining technique namely Decision Tree , Rule Based Classifiers with Compare to Linear Regression as a standard statistical data mining method to predict graft survival period of kidney transplants over a 5-year horizon. The dataset was compiled from the Urology and Nephrology Center (UNC), Mansoura, Egypt. classifiers were developed using the Weka machine learning software workbench by applying Rule Based Classifiers (M5Rules),Decision Tree Classifiers (REPTree) and Linear Regression. Further from Experimental Results, it has been found that Rule Based classifiers and Decision Tree are providing improved Accuracy and interpretable models compared to other Classifier.

Keywords
Data Mining, Renal Transplanat, Classification, Rule Based,Decision Tree, Linear Regression.

I. Introduction

In March 1976, the first renal transplant in Egypt was carried out at the Department of Urology,University of Mansoura. A mother donated one of her kidneys to her daughter who was suffering from end-stage renal disease secondary to chronic pyelonephritis. Armed only with azathioprine and corticosteroids, the operative procedure and the functional outcome were very successful. Atypical example of beginner's luck. Following a very slow start, the number of procedures increased gradually until it has currently reached a rate exceeding 80 cases every year[1].

The importance of having a possibility to predict the outcome after renal transplant is helpful for decision makers to help better manage the overall renal transplant process starting with who should get the renal and allow the choice of the best possible kidney donor and the optimum immunosuppressive therapy for a given patient and this will not only extend the longevity and quality of life for the recipient patient but also reduce medical expenses and increase the access to donor kidneys by reducing the need for multiple kidney transplants in one patient [2,3].

Several prediction methods have been focused upon the use of standard statistical models to predict the outcome of renal transplantation [3-6]. machine learning algorithm applications are widely used in medical fields and in in nephrology "especially" kidney transplant with good results [7-15] comparable" outcome" to traditional statistical tools[16-21].

In this paper we compare the performances and features of data mining technique namely Decision Tree and Rule Based Classifiers with Compare to Linear Regression as a standard statistical data mining method to predict 5-year graft survival of livingdonor kidney transplants. using the patient profile information prior to the transplant process. with the challenge being to select the right kidney from the available kidney donors for a particular patient in order to maximize the chances for the successful transplantation.

The rest of the paper is organized as follows. Related Redearch discussed in Section 2; and in Section 3- Methodology for our proposed work has been detailed; Section 4 outlines the Results and discussion and Section 5
illustrates Conclusions and future work.

II. Related Research

Several studies have been focused on kidney Transplanation

In medical fields There are Several studies have been focused on kidney Transplanation [table1].These studies have applied different Machine Learning Methods to the given problem and have achieved higher prediction accuracies rangingfrom 62% or highe.

Maha Fouad,et al. [22], 1900 patient data obtained from urology and nephrology center (UNC), Mansoura, Egypt, from march 1976 and june 2007. compare the performances and features of data mining technique namely decision tree , rule based classifiers with compare to logistic regression as a standard statistical data mining method to predict the outcome of kidney transplants over a 5-year horizon and found that the highest accuracy is (76.89%) Belongs to RuleBased classifier (JRIP).

Jiakai Li,et al.[23] using the University of Toledo Medical Center (UTMC) patient data as reported to United Network Organ Sharing (UNOS) and had 1228 patient records for the period covering 1987 through 2009. To Predict renal transplantation graft status and graft survival period using Bayes net classifiers , Two separate classifiers were induced from the data set, one to predict the status of the graft as either failed or living, and a second classifier to predict the graft survival period. prediction accuracy of 97.8% and true positive values of 0.967 and 0.988 for the living and failed classes, respectively. The second classifier to predict the graft survival period yielded a prediction accuracy of 68.2% and a true positive rate of 0.85 for the class representing those instances with kidneys failing during the first year, results indicated that it is feasible to develop a successful Bayesian belief network classifier for prediction of graft status, but not the graft survival period, using the information in UNOS database.

Akl A, et al. [24], 1900 patient data obtained from Urology and

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

Nephrology Center (unc), Mansoura, Egypt, From March 1976 and June 2007. To predict5-year graft survival of livingdonor kidney transplantation . comparing two potential methods—an artificial neural network (ANN) and a scoring nomogram calibrated from Cox regression coefficients. The ANNs sensitivity was 88.43 %, specificity was 73.26 %, and its predictions was 16% significantly more accurate than the Cox regression-based nomogram area under ROC curve was 88%. The Cox regression-based nomogram sensitivity was 61.84% with 74.9% specificity and area under ROC curve was 72%. the predictive accuracy of the ANNs prognostic model was superior to that of the nomogram in predicting 5-year graft survival.

J.-H. Ahn et al. [25], using the publicly-available data from the United Network for Organ Sharing UNOS with 35,366 obtained from records for kidney-transplants performed between 1987 and 1991. applied the Bayesian belief network to a large UNOS dataset to develop a predictor for renal graft survival period. The model was developed using a supervised, machine-learning approach, called the Advanced Pattern Recognition and Identification (APRI) system. The APRI system builds the Bayesian network The model was used to predict one-year graft survival rates. They illustrated the model's prediction for two hypothetical kidney-transplant patients. Patient A who is younger, never had a prior transplant, had fewer HLA mismatches, and a lower peak panel reactive antibody level was compared to those of patient B. Because of these favorable health characteristics, patient A had a much higher average predicted graft survival rate (91.2%) than patient B (78.4%). Finally, they claimed the performance in predicting 1-year graft survival rates showed promise for providing valid information to better allocate such scarce resources as transplant organs..

D. Lofaro et al [7], sample of 80 consecutive renal transplants performed between January 1996 and February 2003 including 52 male and 28 female Caucasians of Overall average age (41.6 ± 12.6) years (range=18±63 years) at time of transplantation. Patient follow-up was 60 months (mean = 55.20 ± 12.74). Researchers have shown two classification trees to predict chronic allograft nephropathy (CAN),(no CAN) through an evaluation of routine blood and urine tests. Classification trees based on the C 4.8 algorithm were used to predict CAN development starting from patient features at transplantation and biochemical test at 6-month follow-up. The first tree model (CAN) in the validation set showed a sensitivity of 62.5%, a false-positive rate of 7.2%, and an area under ROC curve of 0.847 (95% confidence interval [CI] 0.749–0.945) and reports the second tree model (no CAN) that showed a sensitivity of 81.3%, a false-positive rate of 25%, and an area under ROC curve of 0.824 (95% CI 0.713–0.934) in the validation set. Identification models have predicted the onset of multifactorial, complex pathology, like CAN. The use of classification trees represent a valid alternative to traditional statistical models, especially for the evaluation of interactions of risk factors.

Fariba ,et al[26], they conducted an experiment on graft outcomes prediction using a kidney transplant datasetbut Not determined. predict the outcome of kidney transplants over a 2-year horizon. compared a widely used ANN approach known as Multi-layer Perceptron (MLP) networks with logistic regression, it has been found that ANN coupled with bagging is an effective data mining method for predicting kidney graft outcomes. and confirmedthat different techniques can potentially be integrated to obtain a better prediction. and proved that a limitation of the ANN approach is that the way predictions are produced is not obvious.

Lasserre Jet al.[27], data comprise 707 transplantations performed at Charité- Universitätsmedizin Berlin (Campus Virchow-Klinikum) between 1998 and 2008. to predict the estimated glomerular filtration rate (eGFR) of the recipient 1 year after transplantation from donore-recipient data using f linear regression (LR) and support vector machines with a Gaussian kernel (G-SVMs) ,neural networks (NNs) and random forests (RFs). the authors obtained a Pearson correlation coefficient between predicted and real eGFR (COR) of 0.48. The best model for the dataset was a Gaussian support vector machine with recursive feature elimination on the more inclusive dataset.

III. Material and Methods

A. Proposed Methodology

The proposed methodology and overall framework of the study for every machine learning techniques (see Figure 1).

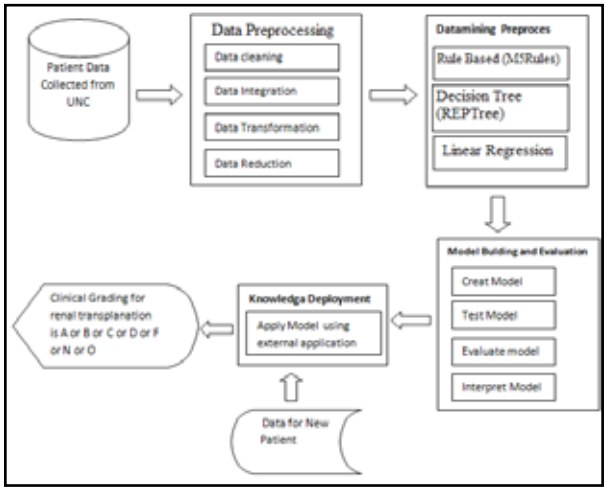

Fig1: Proposed methodology and overall framework of the study

B. Data Mining Process

The following steps of data mining were carried out in order to effectively apply data mining [28],[29].

1. Application Domain

Experiment was conducted in the Urology and Nephrology Center (UNC), Mansoura, Egypt.

2. Data set selection

Data were selected on the basis of the recommendations of the expert doctor.Between March 1976 and June 2007, 1900 consecutive living-donor renal transplants were performed in the Urology and Nephrology Center, Mansoura, Egypt. For recipients, our exclusion criteria included sensitization with a positive lymphocytetotoxic crossmatch, recent malignancy, addiction, psychiatric disorders, type I diabetes mellitus, and significant extra renal organ failure (pulmonary, hepatic, and cardiac). Absolute contraindications to donation included active infections, diabetes, any renal function impairment, arterial hypertension, and positive serology for hepatitis B virus or hepatitis C virus. There were 1564 related donors and 336 unrelated donors, including 118 spouses.Graft loss was defined as graft failure or patient's loss. The study applies to transplants that have complete records and have survived beyond 3months posttransplantation [24] figure (2) Displays examples

of Attributes or variables used in ourexperiment. Common fields were used in expriment according to Doctor's recommendations . The time-to-failure for the transplanted kidney is variable of interest for observation or prediction for this study, so the diff erence between date of transplanation and date of last follow up for the patient was added to track the periods of the survival time after the renal transplantation under a field named "graft survival period".

3. Data cleaning and preprocessing

Data pre-processing is the important step in data mining because In "real world" database, will be The incomplete, inconsistent and noisy data. Therefore preprocessing is a very important stage [30]. In this experimrnt Noisy and inconsistent records were removed, redundancies variables were removed to prevent errors in the dataset , discretization by transform some variables from nominal value to numeric values and others from numeric value to numeric , Normalization of the numerical values into the interval , for missing values the data were used as it is besause the used algorithm support missing values[31].

4. Data formatting:

Experiment done using WEKA (version 3.6.10) which is a suite of software learning machine written in Java and was developed at the University of Waikato (New Zealand). It is free software available under the GNU General Public License. In Weka System two data file formats are used, CSV(Comma SeparatedValue) or an ARFF(attribute relation file format)file is an ASCII text file that describes a list of instances sharing a set of attributes. so data were converted to a standard format CSV and Arff[32] . these are snapshot for training set used in our experiment in ARFF format (see Figure 2).

```
@relation 'patient_data-weka.filters.unsupervised
.attribute.Remove-R1-weka.filters.unsupervised.
attribute.Remove-R47-49-weka.filters.unsupervised.
attribute.Remove-R15,43'
@attribute age.reci numeric
@attribute sex.reci numeric
@attribute or.kid.d numeric
@attribute consang numeric
@attribute rec.b.g numeric
@attribute blood.gp numeric
@attribute number.o numeric
@attribute hypr_pre numeric
@attribute dial_pre numeric
@attribute age.donr numeric
@attribute sex.donr numeric
@attribute don.b.g numeric
@attribute hla numeric
@attribute dr numeric
@attribute tran.rec numeric
@attribute isc.time numeric
@attribute tim_diu numeric
@attribute num.rena numeric
@attribute ren.vein numeric
@attribute p.urin.r numeric
@attribute prim.imm numeric
@attribute secn.imm numeric
@attribute ind.s.im numeric
@attribute tert.imm numeric
@attribute ind.t.im numeric
@attribute num.arej numeric
@attribute num.g.bp numeric
@attribute c_rej_no numeric
@attribute tot.dos1 numeric
@attribute cond.dis numeric
@attribute liv_dial numeric
@attribute ser_cr1y numeric
@attribute clin.gr1 {A,N,I,B,F,C,T,D,O,.}
@attribute ser_cr2y numeric
@attribute clin.gr2 {A,B,N,I,C,D,F,.}
@attribute ser_cr3y numeric
@attribute clin.gr3 {A,B,C,D,I,F,N,O}
@attribute ser_cr4y numeric
@attribute clin.gr4 {A,B,I,C,D,F,N,O}
@attribute ser_cr5y numeric
@attribute clin.gr5 {A,B,I,C,F,D,N,O}
@attribute Graft_surv_per_in_month numeric
@attribute ser_crea numeric
@attribute clin_gra {F,N,I,B,A,C,D,T,O}
@data24,2,6,1,4,1,8,1,1,42,2,4,3,2,1,50,1,1,1,1,1,0,8,0,8,0,0
,0,6.5,2,1,?,?,?,?,?,?,?,?,?,?,?,84,2.2,F,29,2,88,2,2,1,8,1,1,30,2
,2,8,8,1,50,2,1,1,1,1,0,8,0,8,1,0,1,6.4,2,1,?,?,?,?,?,?,?,?,?,?,5
,?,F
```

Fig 2: data in ARRFF formate

5. Choosing the Function (Method) of Data Mining:

Classification is known assigning an unknown object to a predefined class after examining its characteristics . in machine learning Classification is considered as supervised learning. In classification learning, classified examples are presented with the learning scheme so , it is expected to learn a way of classifying unseen examples. [33] in our expriment classification method were applied using different datamining algorithms As Rule Based Classifiers (M5Rules),Decision Tree Classifiers (REPTree) and Linear Regression. using WEKA (version 3.6.10) .

6. Choosing the Data Mining Algorithm

i. Rule Based Classifiers

A Rule Based Classifier is considered as a classification technique that use logic propositional formulas in disjunctive or conjunctive normal form ("if then rules") for classifying the given records, this classification technique is also called ruled based [34]. Rule Based Classifiers Produce Descriptive Models ,and easy to interpret ,Especially in medical field through providing the medical doctor with a compact view of the analyzed data .We Applied the below as an examples of Rule Based Classifiers.

M5Rules: Generates a decision list for regression problems using separate-and-conquer. In each iteration it builds a model tree using M5 and makes the "best" leaf into a rule.[m5rule]

By Appling "weka.classifiers.rules.M5Rules"

ii. Classification Trees

Classification Trees, i.e. Decision tree is a classification method commonly used in data mining [35]. Decision Trees Are used to create a model that predicts the value of a target variable based on several input variable. In our expriment WEKA was used REPTree to build the tree model and perform the classification analysis.

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

REPTree: Fast decision tree learner. Builds a decision/regression tree using information gain/variance and prunes it using reduced-error pruning (with backfitting). Only sorts values for numeric attributes once. Missing values are dealt with by splitting the corresponding instances into pieces (i.e. as in C4.5).
By Applying " weka.classifiers.trees.REPTree".

iii. Linear Regression

In linear regression, data are modeled using linear predictor functions, and unknown model parameters are estimated from the data. Such models are called linear models. Most commonly, linear regression refers to a model in which the conditional mean of y given the value of X is an affine function of X. Less commonly, linear regression could refer to a model in which the median, or some other quantile of the conditional distribution of ygiven X is expressed as a linear function of X. Like all forms of regression analysis, linear regression focuses on the conditional probability distribution of y given X, rather than on the joint probability distribution of y and X, which is the domain of multivariate analysis[36].
By Appling "weka.classifiers.functions.LinearRegression " Where it is Class for using linear regression for prediction . Uses the Akaike criterion for model selection, and is able to deal with weighted instances.

7. Data Mining (Pattern Extraction)

when applying the Predifined algorithms we obtained different models algorithm that can be used to classify, predict, or rule out new clinical cases.
M5Rules Classifier model when The M5Rules algorithm applied as seen in the following figure (3).

```
M5 pruned model rules
(using smoothed linear models) :
Number of Rules : 11
Rule: 1
IF
      clin.gr5=F > 0.008
      clin.gr3=N <= 0.003
THEN
Graft_surv_per_in_month =
      -0.1464 * sex.reci
      - 0.0023 * or.kid.d
      - 1.2662 * consang
      - 1.3278 * number.o
      - 0.2282 * age.donr
      + 4.1312 * sex.donr
      - 0.0503 * dr
      - 0.0121 * isc.time
      - 0.1252 * tim_diu
      - 0.2679 * p.urin.r
      - 0.1154 * prim.imm
      - 0.1402 * secn.imm
      - 0.0899 * ind.s.im
      + 0.8614 * tert.imm
      + 0.1076 * ind.t.im
      - 0.0533 * c_rej_no
      + 1.8598 * tot.dos1
      + 0.1018 * liv_dial
      - 4.5524 * ser_cr1y
      + 1.3333 * clin.gr1=A
      - 1.7144 * clin.gr1=N
    - 0.7744 * clin.gr1=I
      + 0.9097 * clin.gr1=B
      + 0.7914 * clin.gr1=D
      - 0.9303 * clin.gr1=.
      - 0.1363 * ser_cr2y
      + 0.4026 * clin.gr2=A
      - 1.4275 * clin.gr2=I
      + 1.7249 * ser_cr3y
      - 1.1028 * clin.gr3=N
      - 0.0693 * ser_cr4y
      - 8.0976 * clin.gr4=A
      - 0.389 * clin.gr4=I
      - 0.8038 * clin.gr4=N
      - 0.3071 * clin.gr5=A
      + 0.588 * clin.gr5=B
      + 0.2258 * clin.gr5=C
      + 0.4446 * clin.gr5=F
      + 2.5565 * ser_crea
      - 0.6385 * clin_gra=N
      + 11.8478 * clin_gra=B
      + 11.2158 * clin_gra=A
      + 49.1237 [316/30.352%]
```

Fig. (3) Classifier model when The M5Rules algorithm applied .

REPTree Classifier model when The REPTree algorithm applied as seen in the following figure (4) and figure (5).

```
REPTree

============
prim.imm < 5.5
|  clin.gr1=A < 0.5
|  |  clin.gr1=B < 0.5 : 22.95 (93.91/1041.05) [40.9/690.9]
|  |  clin.gr1=B >= 0.5
|  |  |  tert.imm < 0.5
|  |  |  |  ser_cr2y < 2.15
|  |  |  |  |  tot.dos1 < 3.75
|  |  |  |  |  |  ind.s.im < 4.5 : 82.37 (11.17/734.42)
[4.17/1043.97]
|  |  |  |  |  |  ind.s.im >= 4.5 : 44.11 (17.75/856.42)
[7.96/1491.19]
|  |  |  |  |  tot.dos1 >= 3.75
|  |  |  |  |  |  clin.gr2=I < 0.5 : 101.86 (77.59/2181.45)
[36.57/2737.96]
|  |  |  |  |  |  clin.gr2=I >= 0.5 : 28.37 (5.58/300.95)
[1.95/435.19]
|  |  |  |  ser_cr2y >= 2.15
|  |  |  |  |  tot.dos1 < 3.75
|  |  |  |  |  |  liv_dial < 0.5 : 15.98 (4.99/127.09) [2.13/73.81]
|  |  |  |  |  |  liv_dial >= 0.5 : 53.6 (4.39/280.28) [9.21/447.43]
|  |  |  |  |  tot.dos1 >= 3.75 : 63.94 (61.61/840.22)
[14.23/3575.53]
|  |  |  tert.imm >= 0.5
|  |  |  |  prim.imm < 4.5
|  |  |  |  |  clin_gra=I < 0.5
|  |  |  |  |  |  age.donr < 23.5 : 64.5 (2/1122.25) [0/0]
|  |  |  |  |  |  age.donr >= 23.5
|  |  |  |  |  |  |  liv_dial < 0.5 : 216.01 (7.27/971.64) [1/1.27]
|  |  |  |  |  |  |  liv_dial >= 0.5 : 155.65 (13.55/1076.43)
```

```
[5/168.51]
| | | | | clin_gra=I >= 0.5
| | | | | | ser_cr1y < 1.6 : 193.07 (2.04/1731.16) [0/0]
| | | | | | ser_cr1y >= 1.6
| | | | | | | clin.gr2=I < 0.5 : 77.2 (9.19/304.36) [7/959.37]
| | | | | | | clin.gr2=I >= 0.5 : 41.77 (2.04/28.7) [0/0]
| | | | prim.imm >= 4.5
| | | | | ser_cr3y < 2.5
| | | | | | isc.time < 38.5 : 117.31 (7/1064.41) [6/1757.14]
| | | | | | isc.time >= 38.5 : 84.16 (22/844.42)
[8.27/3274.16]
| | | | | ser_cr3y >= 2.5 : 52.83 (9.28/327.51) [5.73/937.95]
| clin.gr1=A >= 0.5
| | prim.imm < 4.5
| | | clin.gr2=A < 0.5 : 76.6 (34.26/2236.69) [15.64/1984.18]
| | | clin.gr2=A >= 0.5
| | | | clin_gra=B < 0.5
| | | | | clin_gra=A < 0.5
| | | | | | clin.gr5=A < 0.5
| | | | | | | clin_gra=N < 0.5
| | | | | | | | ser_cr4y < 3.2 : 120.19 (20.75/1312.99)
[6.67/768.9]
| | | | | | | | ser_cr4y >= 3.2 : 66.65 (4.61/258.78)
[1.15/816.12]
| | | | | | | clin_gra=N >= 0.5 : 69.96 (9.24/1127.02)
[5.52/1906.79]
| | | | | | clin.gr5=A >= 0.5 : 147.71 (65.5/5209.02)
[30.43/5062.69]
| | | | | clin_gra=A >= 0.5
| | | | | | p.urin.r < 1.5
| | | | | | | number.o < 5.5 : 262.75 (4/73.19) [4/1648.69]
| | | | | | | number.o >= 5.5 : 230.81 (4/26.19)
[1.82/31702.84]
| | | | | | p.urin.r >= 1.5 : 188.85 (22/1946.32)
[19/2693.38]
| | | | clin_gra=B >= 0.5
| | | | | p.urin.r < 1.5 : 269.44 (12/625.91) [4/129.51]
| | | | | p.urin.r >= 1.5 : 186.5 (17.03/797.26)
[6.03/6887.66]
| | prim.imm >= 4.5
| | | tot.dos1 < 3.38
| | | | ren.vein < 2.5 : 56.82 (39.88/1094.82) [34.5/1076.11]
| | | | ren.vein >= 2.5 : 92.74 (10/320.4)
[8.63/2527.09]
| | | tot.dos1 >= 3.38
| | | | clin_gra=B < 0.5
| | | | | clin.gr3=A < 0.5
| | | | | | clin.gr3=B < 0.5
| | | | | | | age.donr < 47.5 : 39.42 (5.95/206.85)
[3.38/514.43]
| | | | | | | age.donr >= 47.5 : 69.02 (2.06/357.36)
[2.04/367.99]
| | | | | clin.gr3=B >= 0.5
| | | | | | clin_gra=A < 0.5 : 89.65 (38.92/1093.75)
[25.52/1837.08]
| | | | | | clin_gra=A >= 0.5 : 57.03 (5.72/2368.22)
[4.82/2508.39]
| | | | clin.gr3=A >= 0.5
| | | | | tot.dos1 < 4.05
| | | | | | isc.time < 35.5 : 115.64 (32.25/1371.62)
[22/2367.52]
| | | | | | isc.time >= 35.5 : 71.02 (38.56/1896.1)
[14.5/1750.16]
| | | | | tot.dos1 >= 4.05 : 118.28 (103.17/2989.32)
[54.62/3028.23]
| | | | clin_gra=B >= 0.5
| | | | | ser_cr1y < 1.45
| | | | | | or.kid.d < 12.5 : 140.65 (66.44/1444.58)
[30/1188.14]
| | | | | | or.kid.d >= 12.5
| | | | | | | ser_crea < 2.05 : 101.6 (28.22/667.9)
[15.61/1184.12]
| | | | | | | ser_crea >= 2.05 : 127.04 (19/482.94)
[9/715.3]
| | | | | ser_cr1y >= 1.45 : 37.6 (3.1/64.68) [1.02/234.91]
prim.imm >= 5.5
| prim.imm < 13.5
| | isc.time < 50.5
| | | clin.gr1=. < 0.5
| | | | tim_diu < 1.5
| | | | | num.g.bp < 3.5 : 55.68 (111.98/367.04)
[61.95/500.48]
| | | | | num.g.bp >= 3.5 : 29.05 (4.98/527.56) [3/667.47]
| | | | tim_diu >= 1.5
| | | | | age.donr < 27.5 : 49.33 (2/72.25) [1/6.25]
| | | | | age.donr >= 27.5 : 24.5 (4/70.25) [0/0]
| | | clin.gr1=. >= 0.5 : 5.95 (2.03/0.42) [1.05/2.98]
| | isc.time >= 50.5
| | | clin_gra=F < 0.5
| | | | number.o < 5.5
| | | | | age.reci < 43
| | | | | | tot.dos1 < 3.25 : 58.74 (10/53.44) [9/465.27]
| | | | | | tot.dos1 >= 3.25 : 74.25 (5/58.24) [3/200.83]
| | | | | age.reci >= 43 : 40.2 (4/205.25) [6/562.58]
| | | | number.o >= 5.5
| | | | | prim.imm < 11 : 36.67 (54/467.95) [19/601.83]
| | | | | prim.imm >= 11 : 58.42 (14/65.84) [5/14.48]
| | | clin_gra=F >= 0.5
| | | | don.b.g < 3 : 35.5 (2/56.25) [4/407.25]
| | | | don.b.g >= 3 : 9 (5/30) [3/164.67]
| prim.imm >= 13.5
| | prim.imm < 16
| | | age.reci < 14.5 : 11.79 (17/63.41) [7/79.86]
| | | age.reci >= 14.5 : 27.07 (71/19.02) [29/60.28]
| | prim.imm >= 16
| | | tot.dos1 < 2.23 : 18 (10/1.8) [6/61.67]
| | | tot.dos1 >= 2.23
| | | | num.arej < 0.5 : 2 (8/0.61) [3/1.1]
| | | | num.arej >= 0.5 : 15.83 (3/80.89) [3/19.44]
Size of the tree : 113
```

Fig. (4) : Classifier model when The REPTree algorithm applied

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

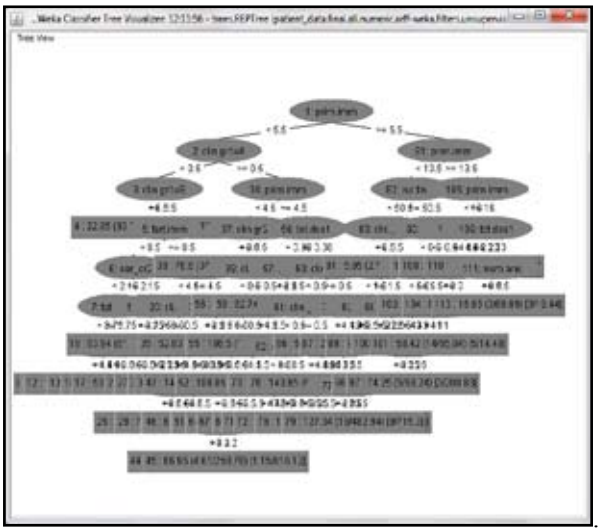

Fig. (5) : Classifier model when The REPTree algorithm applied

Linear Regression Classifier model when The Linear Regression algorithm applied as seen in the following figure (6)

Linear Regression Model

Graft_surv_per_in_month =

-0.1525 * or.kid.d +
-1.6906 * consang +
1.2203 * rec.b.g +
-2.4656 * number.o +
-3.0335 * hypr_pre +
13.6442 * dial_pre +
-0.249 * age.donr +
-3.5235 * dr +
-0.299 * isc.time +
-16.8178 * p.urin.r +
-5.0687 * prim.imm +
-8.313 * secn.imm +
-2.65 * ind.s.im +
9.6799 * tert.imm +
1.0134 * ind.t.im +
-3.0761 * c_rej_no +
3.7899 * tot.dos1 +
-7.2803 * liv_dial +
-82.9367 * clin.gr1=N +
-92.4208 * clin.gr1=I +
-4.6092 * clin.gr1=B +
-83.6576 * clin.gr1=F +
-25.9451 * clin.gr1=C +
-62.3414 * clin.gr1=. +
-3.0954 * ser_cr2y +
17.6229 * clin.gr2=A +
15.5782 * clin.gr2=B +
-56.9342 * clin.gr2=N +
-86.0188 * clin.gr2=I +
16.8404 * clin.gr2=C +
24.3019 * clin.gr2=D +
6.2763 * clin.gr3=A +
-90.1952 * clin.gr3=I +
-30.5676 * clin.gr3=F +
-72.2237 * clin.gr3=N +
-1.9484 * ser_cr4y +
7.8803 * clin.gr4=A +
6.9838 * clin.gr4=B +
-66.1542 * clin.gr4=I +
22.0312 * clin.gr4=D +
-45.2522 * clin.gr4=F +
-42.0851 * clin.gr4=N +
-6.9236 * ser_cr5y +
37.0548 * clin.gr5=A +
16.2309 * clin.gr5=B +
-49.226 * clin.gr5=I +
8.698 * ser_crea +
31.8275 * clin_gra=F +
41.873 * clin_gra=I +
32.3311 * clin_gra=B +
20.4053 * clin_gra=A +
31.3657 * clin_gra=C +
46.2493 * clin_gra=D +
-51.4527 * clin_gra=T +
119.884

Fig.(6) : Classifier model when The Linear Regression algorithm applied

8. Evaluation and Interpretation

Validation

a 10 fold cross validation is used ,the training data is divided into 10 differentparts of equal size. Then one tenth of the instances present in the training set are usedfor testing and the remaining nine tenth for the training.Once the first round of validation is completed, another subset of equal size is used fortesting, and the remaining 90% of the instances used for training as before. The process is iterated 10 times to ensure the all instances become part of the trainingand test set.At the end, the recorded measures are averaged. The number of false positive, false negative, true positive and true negative classifications is simply accumulated across the 10 runs.

IV. Experimental Result and Discussion.

To measure and investigate the performance on the selected classification method and algorithms namely Rule Based Classifiers (M5Rules) ,Decision Tree Classifiers (REPTree) and Linear Regression as a standard statistical data mining method . Table 1 mainly summarizes Performance measures for Different Classifiers. graphical representations of the simulation result are shown below (see Figures 7,8,9) .

Table (1) Performance measures for Different Classifiers

methods	M5Ruls	REP-Tree	linear re-gression
Correlation coef-ficient	0.8748	0.7373	0.7331
Mean absolute error (MAE)	21.1299	30.3651	32.2746
Root mean squared error (RMSE)	30.599	42.7202	42.9498
Relative absolute error (RSE)	41.45%	59.57%	63.32%
Root relative squared error (RRSE)	48.49%	67.70%	68.06%
Time taken to build model in seconds	238.65	7.16	36.19

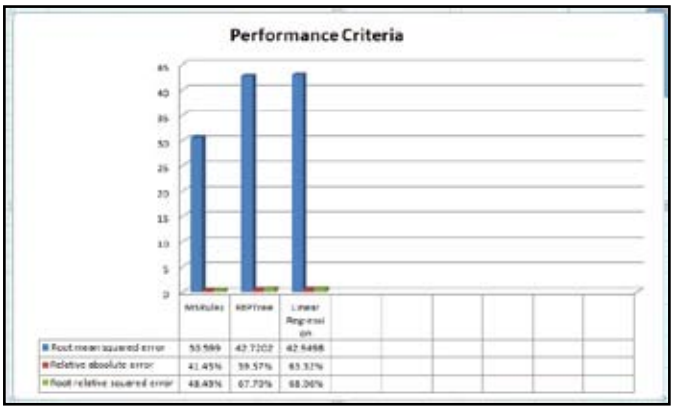

Fig. (7) performance criteria

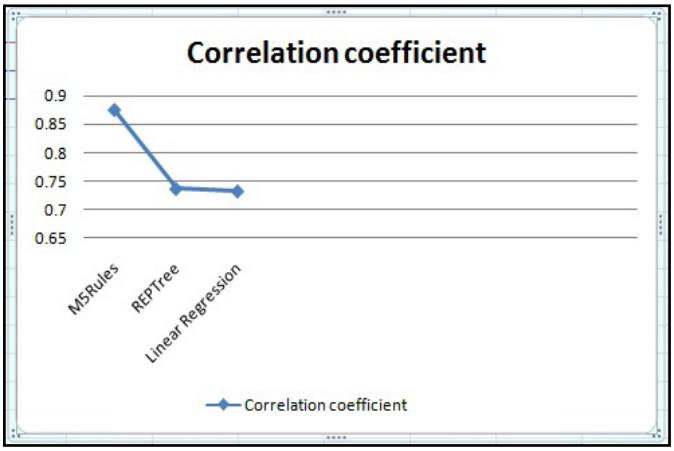

Fig. (8) correlation coefficient

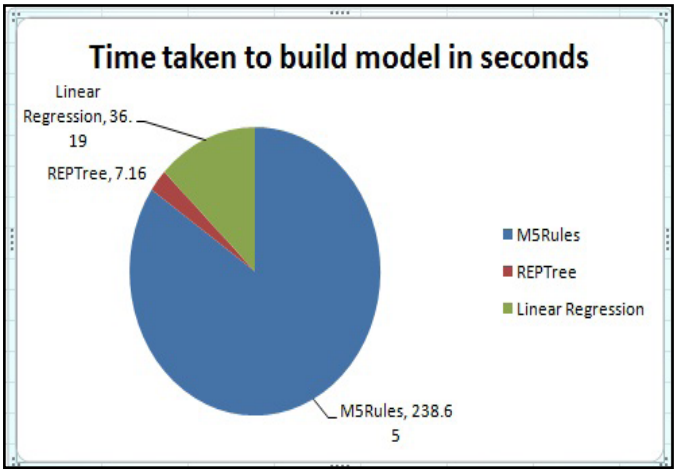

Fig. (9) time taken to build model in seconds

According to the above Figures and Table (1) Performance measures for Different Classifiers , we can clearly see that performance of M5Rules Algorithms is better algorithms regarding values of errors as MAE =21.1299 , RMSE = 30.599 and RRSE =48.49% means it has better prediction but it take more time to perform =238.65 Compared to other algorithms which take less time to perform.

V. Conclusion and Future Work

In this paper, we have discussed the need for data mining in the medical field especially in prediction of kidney transplant outcomes in unc mansoura,egypt. in this context we compared between classification Rule Based Classifiers (M5Rules),Decision Tree Classifiers (REPTree) and Linear Regression as a standard statistical data mining method to predict the outcome of kidney transplants over a 5-year horizon using the patient profile information prior to the transplantation. and found that classification predictive accuracy of rule based classifiers(M5Rules) model was superior other models in predicting 5-year graft survival when run against kidney transplanation dataset obtained from urology and nephrology center, mansoura, egypt.further we have found rule set containing some interesting rules which were easy to interpret and familiar to represent them in spreadsheet were obtained from rule based classifiers and decision tree classifiers.the experimental results also reveal that rule based classifiers and decision tree classifiersare efficient approaches for extraction of patterns from kidney transplanation dataset.

in a future project, we shall implement the rule based classifiers model that was developed in this study in a form of web based application to make it available to estimate survival and prognosticate individual transplant recipients outcomes.

References

[1] *M. A. Ghoneim, M. A. Bakr, N. Hassan et al., "Live-donor renal transplantation at the Urology & Nephrology Center of Mansoura: 1976–1998," Clinical Transplants, pp. 167–178, 2001.*

[2] *Akl A, Mostafa A, Ghoneim M. Prediction of graft survival of living donor kidney transplantation: Nomograms or Artificial Neural Networks. Transplantation 2008; 86: 1401.*

[3] *Poli F, Scalamogna M, Cardillo M, et al. An algorithm for cadaver kidney allocation based on a multivariate analysis of factors impacting on cadaver kidney graft survival and function. Transpl Int 2000; 13(suppl 1): S259.*

[4] *Akl A, Mostafa A, Ghoneim MA. Nomogram that predicts graft survival probability following living-donor kidney transplant. Exp Clin Transplant 2008; 6: 30.*

[5] *Grossberg JA, Reinert SE, Monaco AP, Gohh R, Morrissey PE. Utility of a mathematical nomogram to predict delayed graft function: a single-center experience. Transplantation. 2006;81(2):155-159.*

[6] *Kattan MW. Comparison of Cox regression with other methods for determining prediction models and nomograms. J Urol. 2003 Dec;170(6 Pt 2):S6-9; discussion S10*

[7] *Lofaro D., Maestripieri S., Greco R., Papalia T., Mancuso D., Conforti D., and Bonofiglio R." Prediction of Chronic Allograft Nephropathy Using Classification Trees"© 2010 by Elsevier Inc. All rights reserved ,360 Park Avenue South, New York, NY 10010-1710, Transplantation Proceedings, 42, 1130–1133 (2010)*

[8] *Goto M, Kawamura T, Wakai K, et al: Risk stratification for progression of IgA nephropathy using a decision tree induction algorithm. Nephrol Dial Transplant 24:1242, 2009*

[9] *Binongo JN, Taylor A, Hill AN, et al: Use of classification and regression trees in diuresis renography. Acad Radiol 14:306, 2007*

[10] *Ingram PR, Lye DC, Tambyah PA, et al: Risk factors for nephrotoxicity associated with continuous vancomycin infusion in outpatient parenteral antibiotic therapy. J Antimicrob Chemother 62:168, 2008*

[11] *Jonisch AI, Rubinowitz AN, Mutalik PG, et al: Can highattenuation renal cysts be differentiated from renal cell carcinoma at unenhanced CT? Radiology 243:445, 2007*

[12] *Santori G, Fontana I, Valente U: Application of an artificial*

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

neural network model to predict delayed decrease of serum creatinine in pediatric patients after kidney transplantation. Transplant Proc 39:1813, 2007

[13] Brier ME, Ray PC, Klein J: Prediction of delayed renal allograft function using an artificial neural network. Nephrol Dial Transplant 18:2655, 2003

[14] Fritsche L, Hoerstrup J, Budde K, et al: Accurate prediction of kidney allograft outcome based on creatinine course in the first 6 months posttransplant. Transplant Proc 37:731, 2005

[15] Jahnukainen T, Malehorn D, Sun M, et al: Proteomic analysis of urine in kidney transplant patients with BK virus nephropathy. J Am Soc Nephrol 17:3248, 2006

[16] Akl A, Amani MI, Ghoneim M: Prediction of graft survival of living donor kidney transplantation: nomograms or artificial neural networks? Transplantation 86:1401, 2008

[17] Song X, Mitnitski A, Cox J, et al: Comparison of machine learning techniques with classical statistical models in predicting health outcomes. Stud Health Technol Inform 107:736, 2004

[18] Kattan MW: Comparison of Cox regression with other methods for determining prediction models and nomograms. J Urol 170:S6, 2003

[19] Austin PC: A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting AMI mortality. Stat Med 26:2937, 2007

[20] Terrin N, Schmid CH, Griffith JL, et al: External validity of predictive models: a comparison of logistic regression, classification trees, and neural networks. J Clin Epidemiol 56:721, 2003

[21] Eftekhar B, Mohammad K, Ardebili HE, et al: Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. BMC Med Inform Decis Mak 5:3, 2005

[22] Maha Fouad, Dr.Mahmoud M. Abd ellatif, Prof.Mohamed Hagag, Dr.Ahmed Akl," Prediction Of Long Term Livin Donor Kidney Graft Outcome: Comparison Between Different Machine Learning Methods" international journal of computers & technology vol. 14, no. 2,2015.

[23] Jiakai Li, Gursel Serpen, Steven Selman, Matt Franchetti, Mike Riesen and Cynthia Schneider, "Bayes Net Classifiers for Prediction of Renal Graft Status and Survival Period" International Science Index Vol:4, No:3, 2010

[24] akl a, mostafa a, ghoneim m. prediction of graft survival of living donor kidney transplantation: nomograms or artificial neural networks. transplantation .transplantation 2008; 86: 1401.

[25] J.-H. Ahn, J.-W.Kwon and Y.-S. Lee, "Prediction of 1-year Graft Survival Rates in Kidney Transplantation: A Bayesian Network Model," in Proc. INFORMS & KORMS, Seoul, Korea, 2000, pp. 505-513

[26] Shadabi F., Cox R., Sharma D., and Petrovsky N., "Use of Artificial Neural Networks in the Prediction of Kidney Transplant Outcomes," Lecture Notes in Artificial Intelligence, Vol. 3215, pp. 566-572, 2004.

[27] Lasserre J, Arnold S, Vingron M, Reinke P, Hinrichs C" Predicting the outcome of renal transplantation".J Am Med Inform Assoc. 2012 Mar-Apr; 19(2):255-62. Epub 2011 Aug 28.

[28] Fayyad U, Piatetsky-Shapiro G, Smyth P. Knowledge Discovery and Data Mining: Towards a Unifying Framework. In: Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining. AAAI Press; 1996. p. 82-8.

[29] Ellatif, A. & Mohamed, M. Association rules technique to evaluate software users satisfaction.http://ssrn.com/abstract=1078506, (November 2014)

[30] kotsiantis s., kanellopoulos d., pintelas p., "data preprocessing for supervised leaning", international journal of computer science, 2006, vol 1 n. 2, pp 111–117.

[31] Cole, J. C.. How to deal with missing data. In J. W. Osborne (Ed.), Best practices in quantitative methods(pp. 214–238). Thousand Oaks, CA: Sage. 2008

[32] frank e. and witten i. h., "generating accurate rule sets without global optimization," in in: proc. of the 15th int. conference on machine learning. morgankaufmann, 1998.

[33] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann Publishers Inc., San Francisco,CA, USA, 2nd edition, 2005.

[34] Ahmed Akl, Amani M. Ismail, and Mohamed Ghoneim," Prediction of Graft Survival of Living-Donor Kidney Transplantation: Nomograms or Artificial Neural Networks?" Transplantation • Volume 86, Number 10, November 27, 2008

[35] Rokach, Lior; Maimon, O. (2008). Data mining with decision trees: theory and applications. World Scientific Pub Co Inc. ISBN 978-9812771711.

[36] https://books.google.com.eg/books?id=MjNv6rGv8NIC&pg=PA1&redir_esc=y#v=onepage&q&f=false