

An Efficient Meta Classifier Technique For Membranous Nephropathy Kidney Disease

K.Padmavathi, Dr.A.V.Senthilkumar

Abstract: Kidney disease is the most generally perceived issue of people in the regular world. The filtration of the kidney occurs in the glomerulus. Nearly one fifth of the blood volume that enters the kidney is filtered. Membranous Nephropathy (MN) is the most progressive reason behind the nephrotic issue and it is a model of immune system glomerular sickness. Around 70-80% of patients with essential MN, a counteracting agent called PLA2R antibodies are found in the kidney. This antibody assigns phospholipase A2 receptor and causes the ailment. Prediction plays a key role in identifying the disease. Classification makes the disease diagnosis prediction based on the class labels. The proposed work is to analyze the kidney dataset with classification techniques. The meta classifiers bagging and voting method are considered for the reenactment with the dataset [6]. The performance of the meta classifier algorithm is displayed and it is visible that bagging outperforms the voting method and the results show that classification performs well on the dataset for predicting the Membranous Nephropathy (MN).

Keywords : Membranous Nephropathy, Chronic kidney dataset, meta classifiers, bagging, voting.

1. INTRODUCTION

Kidney disease is the most prevailing problem in today's world. Membranous nephropathy is the outcome of the nephrotic disorder [1]. The human kidney is given the little minor structures called nephrons. When the small nephrons in the kidney get damaged it causes the kidney disease [2]. Membranous Nephropathy is categorized into two types primary MN and secondary MN. Membranous nephropathy (MN) is one among the common capillary disease diagnosed in transplanted kidneys[11]. Filtration barrier of the kidney consists of capillary epithelium cells, capillary basement membrane and extremely specialized epithelial cells called the podocyte. In patients with primary MN, once this filtration barrier is disabled it results in vast loss of proteins in excretion (proteinuria) and it is termed as glomerulonephritis. So that Membranous nephropathy disease is also known as membranous glomerulonephritis(mgn)[15]. MN can recur after kidney transplantation causing proteinuria, allograft dysfunction and graft failure[5]. PLA2R antibodies play a prominent role within the pathological process of primary MN and square measure related to the persistence of nephrotic symptom[3]. The world is moving quickly and the increase in disease is also becoming very fast. In recent technology information mining plays an important role in analyzing and predicting the successful data mining process provided a piece of valuable information for all sectors involved in healthcare-related industries. Classification is one in all the necessary data mining techniques for illness prediction[13]. Predicting membranous nephropathy relies on health parameters like random glucose level, blood serum creatinine level and others[15].

We have analyzed two classifiers vote and bagging method for prediction of this kidney disease. Bagging is a meta algorithm plotted to improve the soundness and accuracy of machine learning algorithms utilized in statistical classification and regression[9]. Bagging constructs n classification trees victimization bootstrap sampling of the training information so combines their predictions to supply a final meta-prediction. A voting algorithm is used for finding the majority of a sequence of components victimization used in linear time and constant house. The voting algorithm, finds a majority element, if there's one: that is, a element that happens repeatedly for over the half the elements of the input[5]. In this proposed method, the meta classifier calculations are connected on the dataset to recognize the best classifier for a forecast of kidney illness for the patient. The dataset is hold on within the Attribute-Relation File Format (.ARFF format) because the information sort of the attributes should be declared [10]. The Classification methodology involves the subsequent steps: a. produce a coaching information dataset. b. verify the class attributes and categories. c. verify useful attributes for classification (Relevance analysis). d. Learn a model victimization coaching examples within the coaching information dataset e. Use the classification model to classify the unknown information. WEKA toolkit 3.8.2 is used for generating the classification techniques and forecast of the result. The system doesn't mechanically classify the attribute as being real or categorical[12].

2. PROPOSED METHOD

A Kidney disorder is one of the challenging diseases in day to day world. The exact factor for identifying the membranous nephropathy disease is unpredictable. But the factors like hypertension, age, case history plays associate degree necessary role in characteristic this disease [11]. Monitoring the PLA2R antibodies plays a really vital role in characteristic the membranous nephropathy disease. Hypothesis methodology is established in our previous work to monitor the PLA2R antibodies. Monitoring the PLA2R antibodies helps us to identify the cure and no cure methods of membranous nephropathy disease[14]. Kaplan meier method is used for survival performance. The empirical analysis is carried out for the patients who continue with the drug therapy and combined drug therapy

-
- ¹ Research Scholar, PG Research and Computer Application, Hindusthan College of Arts and Science, Coimbatore .
 - ¹ padhukarups3@gmail.com
 - ² Director, PG Research and Computer Application, Hindusthan College of Arts and Science, Coimbatore.
 - ² avsenthilkumar@yahoo.com.

with PLA2R anti-body monitoring over a while. Based on that, the survival (cure, no cure) for the membranous nephropathy is determined. The statistics value is calculated and is less than 5. In our proposed work we use the best classification technique meta classifier to predict the disease so that the necessary action can be taken to save the life of a patient. The machine learning algorithm weka 3.8.2 toolkit is used for the performance of this classification technique. Classification techniques classify data based on the training set and class labels. The classification procedure is used to repeatedly make decisions in new situations. In data mining, different classification methods are there to classify data and predict variables. Machine learning algorithmic program meta classifier is employed in our work to predict the illness. The prediction of the illness relies on the accuracy level of the algorithms.

3. METHODOLOGY

The meta classifier method is used in the field of medical diagnosis system to get the better prediction of a disease. There are a different method used in the meta Table I Field information about the kidney dataset

Field Name	Field Description	Field Attribute
age	Age of the patient	Numeric
bp	Blood pressure of the patient	Numeric
sg	Specific Gravity of the patient	Numeric
al	Albumin of the patient	Nominal
su	Sugar level of the patient	Nominal
rbc	Red blood cell count	Nominal
pc	Pus cell of the patient	Nominal
pcc	Pus cells clump of the patient	Nominal
ba	level of bacterial cells	Nominal
bgr	random blood glucose level of the patient	Numeric
bu	level of blood urine	Numeric
sc	creatinine level of the patient	Numeric
sod	level of sodium	Numeric pot
Potassium Level of the patient		Numeric
hemo	Hemo protein level of the patient	Numeric
pcv	Packed cell volume level of the patient	Numeric
wbcc	count of white blood cell	Numeric
rbcc	count of red blood cell count	Numeric
htn	Hyper tension	Nominal
dm	Diabetes mellitus	Nominal
cad	heart disease of the pateint	Nominal
appet	Appetite level	Nominal
pe	Pedel edema	Nominal
anemia	Level of blood	Nominal

Table I describes the various attributes in the kidney dataset. The attributes are represented in numeric and nominal values. In this dataset 12 datasets are considered for numeric values and 13 for nominal values. In the above table, there are totally 400 instances into that information is processed with totally different attributes like glucose,age,bacteria,urea,creatinine,haemoglobin,sodium,potassiuj,cell volume, RBC(red blood cells),pus cells, sugar level of the patient, coronary artery sickness, appetency level, blood pressure.

classification techniques. Bagging and voting algorithms are considered for our proposed work. In classification technique, the whole dataset is divided into predefined classes. So that the classification process is also called supervised learning [8]. The Whole dataset is implemented using the weka tool. To predict the kidney disease UCI repository kidney dataset is considered[16].The dataset is represented in Table I with nominal values and numerical values. The main attribute for predicting the disease is represented in numerical values. The performance level of the algorithmic program is then calculated by the elaborate accuracy of the category followed by the cross-validation outline. To gauge the performance level of the classifier 10-fold cross-validation check is employed. During this technique, the classifier splits the information data into a training sample to train the model and a check set is taken to gauge it. The time taken to create up this model is taken in seconds. The performance level of the algorithmic program is then calculated by the elaborate accuracy of the category followed by the ten fold cross-validation outline.

4. EXPERIMENTAL ANALYSIS AND RESULTS

The experimental comparison of classification algorithms is finished supported the performance based on classification accuracy and execution time. WEKA could be a machine learning workbench that implements associate degree algorithmic program for information pre-processing. Table I dataset is taken into account for experimental comparison. Further, the information is reborn to ARFF (Attribute Relation File Format) format method in WEKA. An ARFF

file is an associate degree computer code document that describes a list of instances sharing a group of attributes.

Table II Representing the cross-validation results

Classifier Methods	Correctly Classified Instance	Incorrectly classified Instance	Time taken in(seconds)
Voting	62.5	37.5	0.5
Bagging	98.75	1.25	0.4

Table II represents the 10-fold cross-validation check. The time taken to make up the model is '0' seconds. During this validation technique, the information is divided into 10 equal size parts wherever nine of those parts are used as a coaching set and therefore one represents the 10-fold cross-validation check used as check data. This methodology continues for ten using a completely different portion as check and training.

Bagging and voting methods are considered for calculating the accuracy levels. Precision, recall, F-measure, Accuracy are the various parameters considered for calculating the accuracy levels. consider a sample of 100 values (x) and we'd wish to get an estimate of the mean of the sample $\text{mean}(x) = 1/100 * \text{sum}(x)$ Improvement of the estimate of the mean exploitation the bootstrap procedure.

1. Generate several (e.g. 1000) random sub-samples of our dataset with replacement.
2. Find the mean of each sub-sample.
3. Find the typical of all of our collected mean and use that as our calculable mean for the information.

Accuracy of voting Algorithm

Input: The value of the chronic kidney dataset.

Precondition: The input document is a text document.

Output: Accuracy of the voting algorithm.

Begin

- 1) Set an element m and a incremental value i.
- 2) Set the incremental value i to be zero.
- 3) For each element, calculate the input sequence x.
- 4) Check the condition loop i.
- 5) Check the condition if whether $i = 0$, then assign $m = x$ and $i = 1$.
- 6) Then calculate the loop
- 7) else check the loop condition if $m = x$,
- 8) Then calculate the increment value of i by $i = i + 1$.
- 9) else
- 10) Calculate the Decrement value of i by $i = i - 1$.
- 11) Return the value of m.

End

RESULTS

Table III Results of Bagging Method

S.no	Parameters	Calculations	Values
1	$\text{Precision} = \frac{TP}{TP + FP}$	$\text{Precision} = \frac{246}{246 + 4}$	0.984
2	$\text{Recall} = \frac{TP}{TP + FN}$	$\text{Recall} = \frac{246}{246 + 1}$	0.997
3	$F - \text{Measure} = \frac{2 * PR}{P + R}$	$F - \text{Measure} = \frac{2 * 0.984 * 0.99}{0.984 + 0.99}$	0.983
4	$\text{Performance Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$	$\text{Performance Accuracy} = \frac{246+149}{246+149+1+4}$	0.9875

--	--	--	--

Table III specifies the performance bagging method. Bagging is performed on weka.classifiers.meta.bagging. The predicted values of positive and negative values are used for the calculations of the parameters. From the

parameters of exactitude precision, recall, f-measure the accuracy level is calculated. The table IV describes the predicted values of true and positive.

Table IV Predicting positive and negative values

Predicted values	Actual Values	
	Positive	Negative
Positive	250	0
Negative	150	0

Table IV is used for predicting the true positive and false negative values. The true positive and false positive rates are taken into the method. Commonly the rates are diagrammatically represented as 1.000,0.999 within the type of a decimal.

Table V Results of voting method

S.No	Parameters	Calculations	Values
1	$Precision = \frac{TP}{TP + FP}$	$Precision = \frac{250}{250 + 1}$	0.625
2	$Recall = \frac{TP}{TP + FN}$	$Recall = \frac{250}{250 + 150}$	0.625
3	$F - Measure = \frac{2 * PR}{P + R}$ $F - Measure = \frac{2 * PR}{P + R}$	$F - Measure = \frac{2 * 1 * 0.625}{1 + 0.625} = \frac{1.25}{1.625}$	0.769
4	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$	$Accuracy = \frac{250 + 0}{250 + 150 + 0 + 0} = 0.625$	0.625

Table V specify the performance bagging method. Bagging is performed on weka.classifiers.meta.bagging. The anticipated values of positive and negative values are used for the calculations of the parameters. From the parameters

of exactitude precision, recall, f-measure the accuracy level of the bagging is calculated. Table V below describes about the predicted values of the voting method.

Table VI Results of the Meta classifiers

Evaluation Criteria	Precision Values	Recall Values	F-Measure Values	Accuracy Values
Voting	0.625	0.625	0.500	0.531
Bagging	0.988	0.988	0.988	0.999

The overall performance of the meta classifier algorithms is listed in Table VI. From the results, it is visible that bagging outperforms the voting method and results show that classification performs well on the dataset for predicting the kidney disease. The time taken to build this method is taken in seconds. The comparison of F-measure is considered to be the lowest with the voting method when compared with the bagging method. Precision is the magnitude of the

most intuitive performance live and it's merely magnitude relation of properly expected observations to the overall observations. Accuracy is that the most intuitive performance live and it's merely a magnitude relation of properly expected observation to the overall observations. The performance of the meta classifier is presented in the graphical comparison.

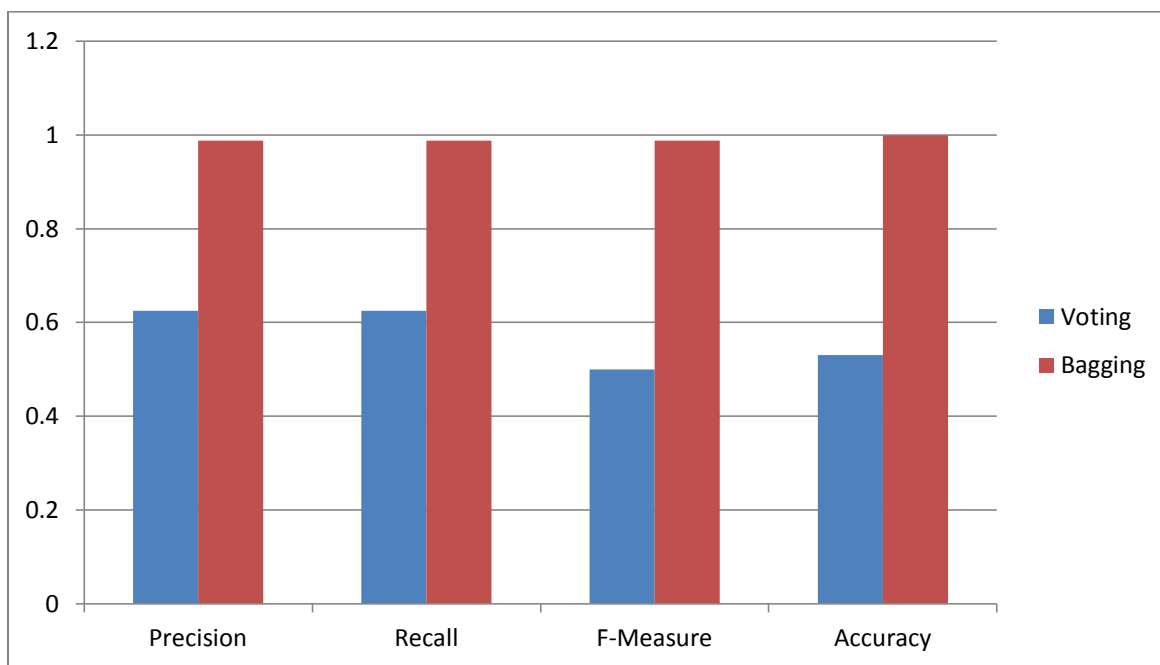


Figure1: Performance of the meta classifiers.

The experimental results show that the classification technique lead to higher accuracy and consumes less time for classification in comparison to the applied statistical methods.

5. CONCLUSION

Membranous nephropathy is a serious implication that may create a life threat when not treated properly. The proposed work is to address the membranous nephropathy disease prediction with data mining technique classification. The bagging and voting algorithms are about to perform their simulation of results. Both classifiers are meta-algorithms. Since meta-classification algorithms are most effective for disease diagnosis, they were utilized. It is evident from the results that, bagging outperforms the voting method which shows a huge improvement in classification cumulative accuracy around 45%. The accuracy of the bagging method performs well on comparing with the voting method. Based on the simulations, the proposed paper will try to focus on the survival analysis interpretation of the samples. The elaborate survival analysis are illustrated and exemplified with the importance of the PLA2R antibodies observation within the treatment of the membranous nephropathy.

6. REFERENCES

- [1] Artur Q.B.da Silva, Taina V. Delaware Sandes- Freitas, Juliana B. Mansur, Jose Osmar Medicina- Pestana, and Gianna Mastroianni - Kirsztajn. "Clinical Presentation, Outcomes, and Treatment of Membranous renal disorder once Transplantation", International Journal of Nephrology, 2018.
- [2] Andrew S. Bomback C. Fervenza, "Membranous Nephropathy: Approaches to Treatment", American Journal of May 2018.
- [3] A S. De Vriese, Richard J. Glasscock, Karl A. Nath, Sanjeev Sethi ,and Fernando C. Fervenza, "A Proposal for a Serology-Based Approach to Membranous nephropathy", www.jasn.org, 2016.
- [4] Bala.S and Kumar, K., 2014, "A Literature Review on Kidney Disease Prediction Using Data Mining Classification Technique," International Journal of Computer Science and Mobile Computing, 3(7), 960-967.
- [5] T.S.Dabade. "Recurrent Idiopathic Membranous Nephropathy After Kidney Transplantation: A Surveillance Biopsy Study", American Journal of Transplantation, 4/29/2008
- [6] Eric Bauer, Ron Kohavi " An Empirical Comparison of Voting Classification Algorithms: Bagging, Booting and Variants, Machine Learning 36, 105–139 (1999) °c 1999 Kluwer Academic Publishers. Manufactured in The Netherlands.
- [7] J. Kittler, M. Hatef, Robert P.W. Duin, J. Matas (1998). On consolidating classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence. 20(3):226-239.
- [8] Ludmila I. Kuncheva (2004). Combining Pattern Classifiers: ways and Algorithms. John Wiley and Sons.
- [9] Mark Hall, Peter Reutemann, "WEKA Knowledge Flow Tutorial", University of Waikato, Jun 2008.
- [10] Manish Kumar Shrivastava, Rohit Miri, Praveenchouskey "Exploring Data mining classification Techniques "International Journal of Engineering Research and Technology. ISSN 2278-0181vo1-2Issue : July 2013
- [11] Rajiv D. Poduval, Michelle A. Josephson, and Basit Javaid, "Treatment of Delaware Novo and recurrent membranous nephropathy in renal transplant Patients", Seminars in Nephrology, vol 23, No 4 (July), 2003: pp 392-399.
- [12] Raj Kumar, Dr. Rajesh Verma, " Classification Algorithms for Data Mining P: A Survey" IJiet Vol. 1 Issue August 2012, ISSN: 2319 – 1058.

- [13] Ramya ,Dr.N. Radha, "Determination of Chronic Kidney Disease Using Machine Learning Algorithms", International Journal of Innovative Research in Computer and Communication Engineering, pp. 812-820 ,volume 4, issue 1, ISSN: 2320-9798, 2016.
- [14] Dr.AV.Senthilkumar, K.Padmavathi "A proposed method for Prediction of Membranous nephropathy Disease "International Journal of Innovative Technology and Exploring Engineering",ISSN:22783075, Volume -8 Issue-10, August 2019.
- [15] Tripti Singh, Brad Astor, Weixiong Zhong, Didier Mandelbrot, Arjang Djamali, and Sarah Panzer , "Kidney Transplant Recipients With Primary Membranous Glomerulonephritis have a higher Risk of Acute Rejection Compared With Other Primary Glomerulonephritis", Kidney transplantation, 2017.
- [16] UCI Machine Learning Repository: Chronic_ Kidney_ Disease Dataset (n.d.).
https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease.