International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075, Volume-8, Issue-12, October 2019

Predicting and Staging Chronic Kidney Disease of Diabetes (Type-2) Patient Using Machine Learning Algorithms

Setu Basak, Md. Mahbub Alam, Aniruddha Rakshit, Ahmed Al Marouf, Anup Majumder

Abstract*: *Mortality because of unending kidney disease increments essentially in recent years. Nowadays, about 422 million patients are suffering from diabetes among them around 30 percent of patients with Type 1 (adolescent beginning) diabetes and around 10 to 40 percent of those with Type 2 (grown-up beginning) diabetes in the end will experience the negative impacts of kidney damage. It is evident, that early detection of Chronic Kidney Disease (CKD) can mitigate the level of damage in the adulthood. In this paper, we have presented a comparative analysis based on the performance of five different algorithms-Naive Bayes (NB), In-stance Based Learning (IBK), Random Forest (RF), Decision Stump (DS) and Decision Tree (J48) for predicting CKD of diabetes patients only by urine test. Among all the algorithms the IBK gives the best result. Our comparison of different algorithms will help people with diabetes to find out if they are having CKD or not.

Keywords:* *Kidney Disease Staging, Cross-Validation, morbidity and mortality, Albuminuria, Proteinuria

I. INTRODUCTION

Chronic Kidney Disease (CKD) is one of the overall general medical issues because of the exorbitant treatment of its end stage and high probability of death [1, 2]. As from bulletin of the World Health Organization (WHO) 2018, Kidney diseases are associated with an estimated 188 million cases of horrifying health expenditure in lower and middle-income countries [3]. In Bangladesh about 18% of the total population are affected with CKD and most of them had type 1 diabetes (39.02%) and type 2 diabetes (41.46%) mellitus. CKD is a situation in which kidneys well-ordered lose their usefulness, making issue waste and abundance liquids gathering in the body and influences the usefulness of the body [4].

As indicated by the overview led by the National establishment found after the analysing of kidney problem, around a large number of people groups were influenced. In this way, the passing rate expanded in renal illness might be because of nourishment item and method for changing in culture the sicknesses spreading more [5]. In some cases, some of the time a man who is influenced by early CKD may not feel unwell or see any prodrome [4]. Again, people with diabetes have high chances of having kidney diseases. For diabetic's patient, if diabetics is not controlled sugar level and Bilirubin of blood goes up. Due to high blood sugar level, blood vessels become narrow and clogged. For this, vessels cannot filter enough blood also albumin passes through these filters in the urine [6]. For nor-mal kidneys, very smaller amount of albumin is found in urine but when albumin level is raises up in the urine that will mean the kidney have become damaged be-cause of diabetics [7]. Diabetes patients whose blood glucose is too high and blood pressure is too high, they are more likely to enhance kidney diseases [8]. For finding malfunction in kidney, some urine and blood tests are mandatory to be taken.

In this paper, we present the comparison of different machine learning algorithm to predict the CKD of diabetes patients and stages of CKD for Bangladeshi perspective. In our study, we have used datasets form the UCI repository.

II. RELATED WORKS

There have been several studies on the prediction of CKD using statistical approach-es, artificial neural networks, image processing, machine learning algorithms and Data mining algorithms. Therefore, we discovered couple of concentrates identified with restorative analysis and repeat utilizing information mining approaches [1-10]. In [4, 9, 10] authors have discussed about the significance of the attribute selection process for predicting CKD. Chetty N. et al. [9] applied the Sequential Minimal Opti-mization (SMO) and k-nearest neighbor classifier (kNN). Nishanth A. et al. [4] pro-posed the Common Spatial Pattern and Linear discriminant analysis classifier for predicting CKD. In [10] Arasu S. D. proposed a novel method Weighted Average Ensemble Learning Imputation (WAELI) for predicting CKD.

In [1, 9, 11] authors have worked on models that predicts the stage of CKD. CKD stages from 1-5 along with the image result for Glomerular filtration rate reading and actions required during each stage are described in [7, 9, 12]. Panwong P. et al. [1] have worked on predicting the CKD

Revised Manuscript Received on October 2, 2019.

* Correspondence Author
Setu Basak*, Department of CSE, Daffodil International University, Dhaka, Bangladesh. Email: setu.cse@diu.edu.bd
Md. Mahbub Alam, Department of CSE, Daffodil International University, Dhaka, Bangladesh. Email: mahbub15-5237@diu.edu.bd
Aniruddha Rakshit, Department of CSE, Daffodil International University, Dhaka, Bangladesh. Email: aniruddha.cse@diu.edu.bd
Ahmed Al Marouf, Department of CSE, Daffodil International University, Dhaka, Bangladesh. Email: marouf.cse@diu.edu.bd
Anup Majumder, Department of CSE, Jahangirnagar University, Dhaka, Bangladesh. Email: anupju.cse20@gmail.com

from stage 3 to 5 using the Hierarchical Deep Convolutional and Synthetic Minority Oversampling Technique for classification & J48 and Naïve Bayes for the prediction. S. P. deng et al. [11] have proposed joined quality articulation strategy with DNA methylation information and created a melded arranged net-work. In [5] M. Edhayadharshini have proposed a method to identify CKD from Computer Tomography (CT) abdominal images where they have used feature selection of the segmented image by Gabor-PHOG features. In [13, 14, 15] authors have discussed the effect of training with the imbalance data. U. N. Dulhare et al. [13] ap-plied Sequential minimal optimization, OneR for classification, Naïve Bayes and Instance Based Learning (IBK) for the prediction task.

III. PROPOSED METHODOLOGY

In this section, we have presented the methodology for predicting the CKD using medical information features. The proposed method that is used for comparison of five different algorithms for our dataset is illustrated in Fig. 1.

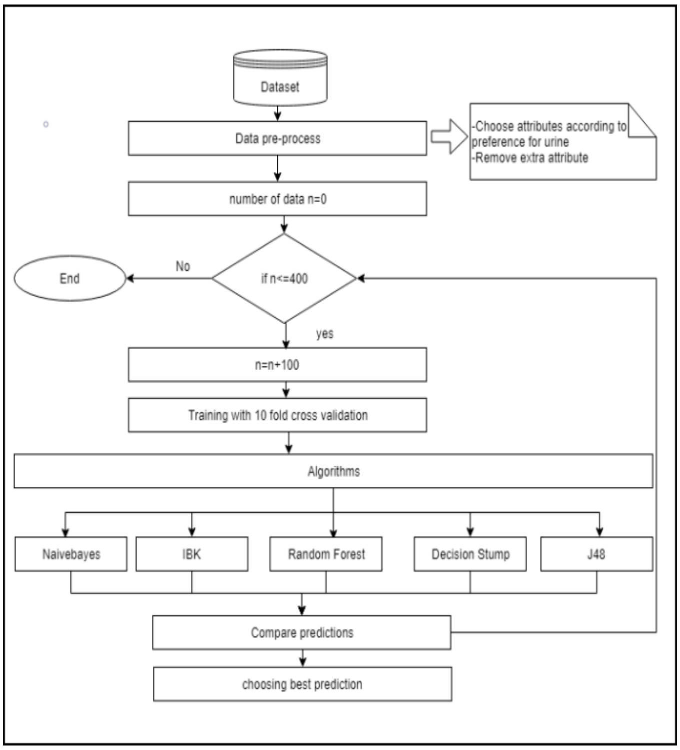

Fig. 1. Flowchart of our working process

Firstly, we processed our dataset. There are 25 attributes in the dataset, among them we used only 5 attributes (Specific Gravity, Albumin, Diabetes Mellitus, Hypertension and Serum Creatinine) because we have focused to measure CKD for diabetes patients by only use of urine test for reducing cost. So, we have removed other attributes. Initially we have taken 100 data from our dataset and trained them using 10-fold cross validation then we have performed five algorithms: NB, IBK, RF, DS and J48 on them. Then, we compare the accuracy of these algorithm and choose the best prediction value. We have iterated this method using n+=100 data each time where n is the previous value. This method continued until the end of the data. And each time we have taken the best predicted value.

IV. DATASET AND ATTRIBUTES

The dataset for this study was obtained from the UCI repository [17]. This dataset is having 24 attributes. This dataset was collected from Apollo Hospitals, Tamilnadu, India. The dataset was created by L. Jerlin Rubini, Alagappa University, Tamilnadu, India.

Dataset consists of attributes such as Age, BP, Specific Gravity, Sugar, Blood Glucose, Albumin, White blood cell count, Red blood cells, Bacteria, Sodium, Potassium, Serum Creatinine, Blood Urea, Hemoglobin, Red Blood Cell Count, Coronary Artery Disease, Packed Cell, Pus Cell clumps, Pus Cell, Diabetes Mellitus, Anemia, Diabetes Mellitus, Appetite, Anemia and class.

V. EXPERIMENTAL RESULT ANALYSIS

We use five different algorithms to compare the best predicting methods of CKD. At first, Table 1 demonstrates the performance matrices of different algorithms. In this paper, we have considered the traditional performance matrices such as precision, recall, F-measure, Accuracy and ROC area. We have considered the attributes that have significant effect on the CKD for type 2 diabetes patients in Fig. 2. The selected attributes are Specific Gravity, Albumin, Diabetes Mellitus, Hypertension and Serum Creatinine.

Table I. Result of Different Classification Algorithms for prediction

Classifiers	**Precision**	**Recall**	**F-measure**	**Accuracy**	**AUC Score**
Naïve Bayes Algorithm	0.958	0.953	0.953	95.25%	.993
IBK	0.983	0.983	**0.983**	**98.25%**	.987
Decision Stump	0.896	0.875	0.877	87.5%	.887
Random Forest	0.972	0.970	0.970	97%	.997
J48	0.978	0.978	0.978	97.75%	.993

Therefore, from the above mentioned results, we can declare that the IBK algorithm gives the best prediction of CKD for diabetes patients about 98.25%.

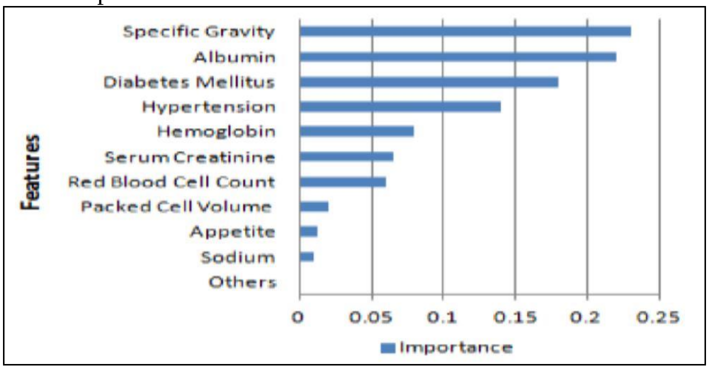

Fig 2. Importance of attributes

The Fig. 3. demonstrates the fluctuations of accuracy rate for the considered algorithms. The

Retrieval Number: paper_id//2019©BEIESP
DOI: 10.35940/ijitee.xxxxx.xxxxxx

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

graph shows that there is a content drop in the False Positive (FP) rate.

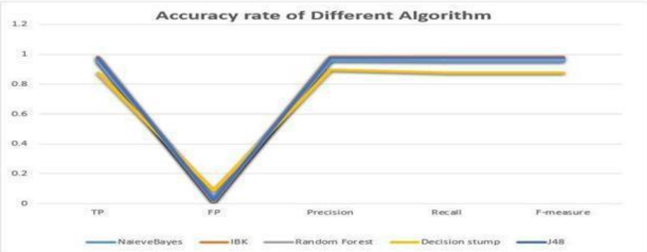

Fig. 3. Line graph showing accuracy of different algorithms

The Fig. 4 demonstrates the accuracy rate of different algorithms at different number of data. From the line graph, we can observe that there is a small fall of accuracy when the number of data is exactly 300 for almost all the algorithms except IBK.

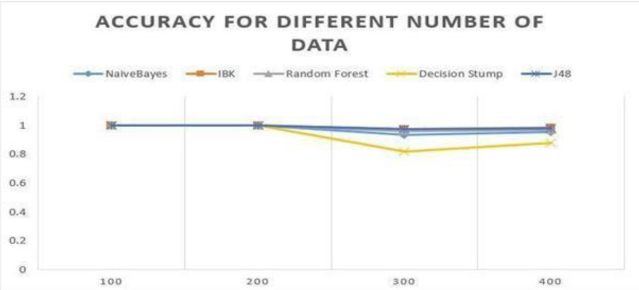

Fig. 4. Line graph of accuracy rate of different algorithms

The Fig. 5 demonstrates the error rate of different algorithms at different number of data. Here, IBK performs best as among all of the error rates, IBK is the least.

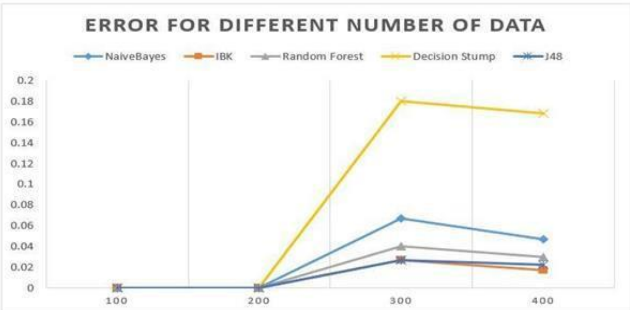

Fig. 5. Line graph of error rate of different algorithms

VI. CKD STAGES

We are using the Albuminuria as for staging the kidney diseases for our dataset. Albuminuria is the presence of protein in urine, ACR (albumin creatinine ratio) is the first method to detect elevated protein that can be done by only urine test which is very much beneficial to a lower- and middle-income country like ours. Steady expanded protein in the urine (two positive tests more than at least 3 months) is the important marker of kidney damage [17].

Albuminuria is an indication of kidney illness and means having excessively albumin in urine. Albumin is a protein found in the blood. Fig 6 represents the amount of protein normally present in healthy and damaged cell [18].

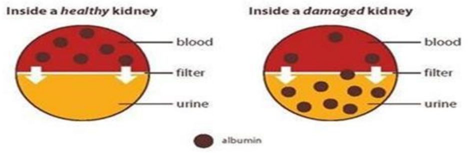

Fig. 6. Healthy Cell Vs Damaged Cell

A sound kidney doesn't give albumin a chance to go from the blood into the urine on the other hand a damaged kidney gives albumin whites a chance to go into the urine illustrated in Fig. 2. The less albumin in the urine, the better. It is additionally called proteinuria. Based on Albumin-to-creatinine ratio we can categorize three stages of CKD patients. Table 2 illustrates the stages of CKD patients for our dataset which is depicted also in Fig. 7.

Table II. Stages of CKD in our Dataset

Stages	**ACR**	**Percentage of**
A1- Normal to mildly increased albuminuria	<30 mg/g	46.75%
A2- Moderately increased albuminuria	30-300 mg/g	5%
A3- Severely increased albuminuria	>300 mg/g	32.5%
None	0	15.75%

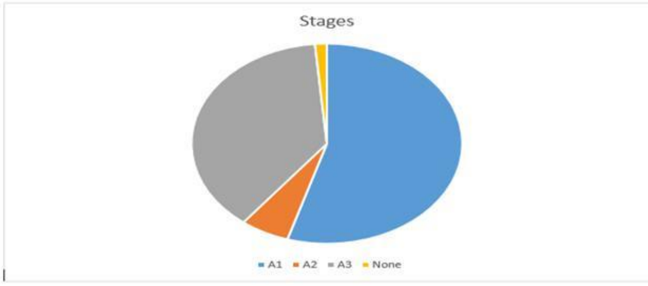

Fig. 7. Pie chart of different CKD stages for our dataset

From both of this table and figure, we can observe that about half our patients of this dataset are suffering from A1(Level 1) CKD. But the number of A3(Level 3) patients are also in danger line. Therefore, CKD can be controlled in A1. But if CKD is not diagnosed before A2 it will be cause of severe kidney damage which leads to death.

VII. CONCLUSION

We applied 5 different algorithms and observed with few data all algorithms give optimal prediction but when the number of data is increasing IBK gives best prediction than others. We also find out the important attributes for CKD where diabetes is one.

In Asia above 60% of the people affected with diabetes are alive [19]. So we can conclude that if Asian people can control their diabetes they will be invaded less with CKD. For future prospects, we are planning to make a system for caring the diabetes patient at early stage by providing appropriate diet plan for individuals.

REFERENCES

1. Panwong, P., Lam-On, N.: Predicting Transitional Interval of Kidney Disease Stages 3 to 5 Using Data Mining Method: Second Asian Conference on Defence Technology (ACDT), pp. 145-150, (2016).
2. Din, U.A.A.S.E., Salem, M.M., Abdulazim, D.O.: Stop chronic kidney disease progression: Time is approaching: PMID: 27152262 World JNephrol. 5(3): 258–273, (2016)
3. Nanditha, A., C.W. Ma, R.C.W., Ramachandran, A., Snehalatha, C., Chan, J.C.N., Chia, K.S., Shaw, J.E., Zimmet, P.Z.: Diabetes in Asia and the Pacific: Implications for the Global Epidemic: Diabetes Care, Volume 39, pp. 472-485, (2016)
4. Nishanth,A., Thiruvaran,T.: Identifying important attributes for early detection of Chronic Kidney Disease: IEEE Reviews in Biomedical Engineering, Volume 11, pp. 208-216,(2017)
5. Edhayadharshini, M., Bhanumathi, V.: Abnormalities Detection in Kidney Using Multithreading Technology: 2017 4th International Conference on Signal Processing, Communications and Networking (ICSCN), pp. 1-6, (2017).
6. Diabetes and Kidney Diseases (Stages 1-4), [Online] https://www.kidney.org/atoz/content/Diabetes-and-Kidney-Disease-Stages1-4 last accessed 2018/9/22
7. Diabetic Kidney Disease, [Online] https://www.niddk.nih.gov/health-information/diabetes/overview/prevenntig-problems/diabetic-kidney-disease Last accessed 2018/7/22.
8. Viberti, G., Wheeldon, N.M.: Microalbuminuria Reduction with Valsartan in Patients with Type 2 Diabetes Mellitus, Circulation (2002), Volume 106, pp. 672-678
9. Chetty, N., Vaisla, K.S., Sudarsan, S.D.: Role of Attributes Selection in Classification of Chronic Kidney Disease Patients: International Conference on Computing, Communication and Security (ICCCS) (2015), pp. 1-6
10. Arasu, S.D., Thirumalaiselvi, D.R.: A Novel Imputation Method for Effective Prediction of Coronary Kidney Disease: 2nd International Conference on Computing and Communications Technologies (ICCCT)(2017), pp. 127-136
11. Deng, S.P., Cao, S., Huang, D.S., Wang, Y.P.: Identifying Stages of Kidney Renal Cell Carcinoma by Combining Gene Expression and DNA Methylation Data IEEE/ACM Transactions on Computational Biology and Bioinformatics (2017), Volume 14, Issue 5, pp. 1147-1153.
12. Diabetic Kidney Disease, [Online] https://patient.info/health/diabetes-mellitus-leaflet/diabetic-kidney-disease. Last accessed: 2018/9/22
13. Dulhare, U.N., Ayesha, M.: Extraction of Action Rules for Chronic Kidney Disease using Naïve Bayes Classifier: IEEE International Conference on Computational Intelligence and Computing Research (2016), pp. 1-5
14. Yildirim, P., Chronic Kidney Disease Prediction on Imbalanced Data by Multilayer Perceptron: IEEE 41st Annual Computer Software and Applications Conference(2017), pp. 193-198
15. Ancy, S., Comelius, K.: A Predictive Model For Kidney Failure E-Health: IEEE 3rd International Conference on Sensing, Signal Processing and Security (ICSSS) (2017), pp. 497-502.
16. Wickramasinghe, M.P.N.M., Perera, D.M., Kahandawaarachchi, K.A.D.C.P.: Dietary Prediction for Patients with Chronic Kidney Disease (CKD) by considering Blood Potassium Level using Machine Learning Algorithms: IEEE Life Sciences Conference (LSC) (2017), pp. 300-303.
17. The renal association, [Online] https://renal.org/informationresources/the-uk-eckd-guide/ckd-stages. Last accessed 2018/10/10.
18. U.S. Department of Veterans Affairs.[Online] https://www.blogs.va.gov/VAntage/36421/four-things-know-chronic-kidney-disease Last accessed 2018/10/10.
19. Arun Nanditha,Ronald C.W. Ma,Ambady Ramachandran,Chamukuttan Snehalatha,Juliana C.N. Chan, Kee Seng Chia,Jonathan E. Shaw, and Paul Z. Zimmet ”Diabetes in Asia and the Pacific:Implications for the Global Epidemic “Diabetes Care 2016 Mar; 39(3): 472-485

AUTHORS PROFILE

Md Mahbub Alam is a researcher who has completed his BSC in CSE degree in 2019 from Daffodil international university. His research interests lies within Data mining, Machine learning and Natural Language Processing.

Aniruddha Rakshit is researcher and working as a Senior Lecturer of Computer Science and Engineering at the Daffodil International University. He received his M.Sc in CSE from Jahangirnagar University in 2017. His recent publication includes “Bengali Spell Checking and Correction Using Edit Distance” and “Recognition of Jute Diseases by Leaf Image Classification Based on Convolutional Neural Network” He received his B.Sc in CSE from the same inistitution in 2014. In 2015, he joined in Daffodil International University as a lecturer in the CSE department. His research interest includes Image Processing, Medical Imaging, Computer Vision, Deep Learning, Data Mining.

Ahmed Al Marouf is full stack researcher and educator, working as a lecturer in Department of Computer Science and Engineering (CSE) of Daffodil International University (DIU), Dhaka, Bangladesh. He also acts as the Technical Lead of DIU HCI (Human Computer Interaction) Research Lab. He has completed his M.Sc. Engg in Computer Science and Engineering (CSE) from Islamic University of Technology (IUT), Gazipur, Bangladesh in 2019. His research interest lies within Computational Social Sciecne, Social Media Analysis, Data Mining and Machine Learning.

Anup Majumder is a research enthusiatic, working as a lecturer in Department of Computer Science and Engineering (CSE) of Jahangirnagar University, Dhaka, Bangladesh. He has completed his M.Sc. in CSE from Jahangirnagar University, Dhaka, Bangladesh in 2015. He received his B. Sc. in CSE in 2014 from the same institution. His research interest lies within Computer vision, Data Mining and Machine Learning.

Setu Basak is a researcher working as a research associate of Computer science and engineering department at Daffodil International University, Dhaka, Bangladesh. He has received his BSC in CSE degree in 2019 from Daffodil international university. His research interests lies in Data mining, Machine learning and Natural Language Processing.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication