ISSN- 2394-5125 VOL 5, ISSUE 06, 2018

CHRONIC KIDNEY DISEASE PREDICTIONS USING MACHINE LEARNING MODELS

ALUGUVELLI KIRAN KUMAR REDDY, AVULA RAJINI DEVI, KORIVI VAMSHEE KRISHNA,
Assistant Professor, Assistant professor, Assistant professor,
Department of CSE,
Samskruti College of Engineering and Technology, Ghatkesar.

ABSTRACT

When it comes to clinical disorders, chronic kidney disease (CKD) is an umbrella term that refers to a wide range of illnesses that deteriorate as kidney function degrades over time. It refers to a wide variety of medical conditions. The term "chronic renal failure" is sometimes used to describe this illness in some circles. Various factors, including genetic abnormalities in the kidneys and systemic illnesses that damage the kidneys, can contribute to chronic kidney disease. Depending on the underlying reason, it might express itself in a variety of ways. Worldwide, the number of people suffering from chronic kidney disease (CKD) is growing year after year, according to the World Health Organization. As defined by the World Health Organization, chronic kidney disease (CKD) is a worldwide public health concern with an increasing incidence and a vast geographic reach that affects individuals all over the world. GFR rises in the presence of renal failure needing dialysis, and it is widely regarded to be the most reliable overall indicator of kidney function in the general population. Heart disease (including high blood pressure and anaemia) and a variety of metabolic problems, to mention a few, are among the additional risk factors for kidney failure. Because of a statistical approach known as 10-fold cross-validation, the algorithms of logistic regression, support vector machines, random forest, and gradient boosting have all been trained and tested on real-world data. According on the F1measure gathered by the classifier after training, the accuracy of the Gradient Boosting classifier is 99.1 percent correct. In addition, we discovered that haemoglobin has a bigger significance for both random forest and gradient boosting in the diagnosis of chronic renal sickness than was previously believed to be the case, which is in direct opposition to previous notions.

1. INTRODUCTION

Long-term kidney disease (CKD) is a serious public health concern that affects individuals all over the world, but it is most widespread in poor and middle-income nations. Chronic kidney disease is caused by a buildup of waste in the kidneys. Eventually, it is caused by a buildup of waste products in the kidneys, which leads to renal failure. As with renal failure, one of the characteristics of chronic kidney disease (CKD) is that the kidney does not function as expected and is unable to filter blood adequately, as is the situation with chronic kidney disease (CKD). Chronical kidney disease (CKD), also known as Chronic Kidney Disease (CKD), is a chronic

ISSN- 2394-5125 VOL 5, ISSUE 06, 2018

kidney disease that affects around 10% of the world's population. A scarcity of inexpensive treatment options causes millions of deaths each year, with the number of deaths among the elderly growing in recent years as a result of a lack of available options. As documented by the International Society of Nephrology's Global Burden of Disease 2010 report, chronic kidney disease (CKD) is a significant cause of death around the world, with the number of fatalities increasing by 82.3 percent over the preceding two decades [1, 2]. Increasing the stakes for patients with chronic kidney disease (CKD). The number of patients who have been diagnosed with end-stage renal disease has also increased significantly in recent years, which is a worrying trend (ESRD). In order to extend the lives of patients who are suffering from this illness, either kidney transplants or dialysis are required. One of the primary goals of their project, which is currently in progress, is to use machine learning algorithms to predict chronic kidney disease in diabetics who are at risk of getting the illness in the future. Following the outcomes of this investigation, the study team examined 600 clinical data from a top diabetic research centre in Chennai, India, for the goal of performing this analysis. The scientists utilised a machine learning programme to analyse the dataset, which featured categorization algorithms such as the decision tree and Naive Bayes, both of which were assessed to be efficient by the scientists who created them. Using the decision tree technique, researchers observed that it outperformed the Nave Bayes algorithm by around nine hundred and ninety percent when compared to the later algorithm. Diagnose and forecast the existence or absence of chronic renal disease in a certain patient (CKD). Each algorithm took into account a total of 25 criteria, and it was concluded that four separate algorithms were preferable to one another in terms of performance. These algorithms were referred to as the Multiclass Decision Forest, Multiclass Decision Jungle, Multiclass Decision Regression, and the Multiclass Neural Network, and each method took into account a total of 14 features in its decision making process. In order to make things easier for you, the findings have been provided as a table. When they finally came to the conclusion that the Multiclass Decision Forest outperformed all other models in their testing, they were relieved to see that the accuracy rate was 99.1.

2. LITERATURE SURVEY

The IEEE Life Sciences Conference (LSC) was held in Sydney, New South Wales, Australia, and the proceedings were published in the IEEE Transactions on Engineering in Medicine and Biology (IEMB) (Vol. 300-303). The IEEE Life Sciences Conference (LSC), which took place this year in Sydney, New South Wales, Australia, was hosted by the city. The research describes in detail how to anticipate dietary needs for individuals with chronic kidney disease (CKD) using machine learning algorithms that take into consideration blood potassium levels, as well as other factors (LSC). Beginning with the following piece of writing is strongly recommended as a starting point for you: Specifically, MPNM Wickramasinghe, D M Perera, and K. A. D C P Kahandawaarachchi state in their paper "Dietary prediction for patients with chronic kidney

ISSN- 2394-5125 VOL 5, ISSUE 06, 2018

disease (CKD) by considering blood potassium level using machine learning algorithms" that they "consider blood potassium level using machine learning algorithms." They also state that they "consider blood potassium level using machine learning algorithms."

A chronic renal disease, as defined by the American Society of Nephrology, is described as one that has caused serious kidney damage and impaired function for more than three months and has persisted for more than three months (CKD). The primary goal of this research project is to determine the best appropriate diet plan for persons with chronic renal disease by analysing test data from patients' medical records and using classification algorithms (CKD). It is anticipated that the project will be completed by the end of the calendar year. Last but not least, the fundamental goal of the study is to aid patients in controlling their conditions through the implementation of a dietary plan that is nutritionally appropriate in terms of nutrients. The final purpose of the study is to determine the most appropriate diet plan via the application of classification algorithms, which will be accomplished through the use of classification algorithms in this study. Following the recommendations of diverse dietary programmes for chronic kidney disease patients based on their blood potassium levels, as predicted by the potassium zone for chronic kidney disease patients, and in accordance with the potassium zone for chronic kidney disease patients, the proposed work pacts are consistent with the potassium zone for chronic kidney disease patients (see Figure 1). Several different algorithms are utilised to determine the outcome of the experiment; and each algorithm is responsible for a certain function. There are a plethora of various multiclass decision-making systems to choose from. Many different types of multiclass decision jungles and forests, multiclass neural networks, Multiclass Logistic Regression, Multiclass Logistic Regression, and Multiclass Logistic Regression are among the most common ones. According to the findings of this study, the Multiclass Decision Forest technique exceeds the other classification algorithms in terms of accuracy, obtaining a rate of 99.17 percent accuracy over a wide range of tests.

The performance of a kernel-based extreme learning machine for the prediction of chronic kidney illness is investigated by H. A. Wibawa, I. Malik, and N. Bahtiar in their study "Evaluation of Kernel-Based Extreme Learning Machine Performance for Prediction of Chronic Kidney Disease." The study was presented at the 2nd International Conference on Informatics and Computational Sciences (ICICoS), which was held in Semarang, Indonesia, in 2018. The article was published in the proceedings of the conference. Due to recent technological advancements in the area of biosciences, in addition to the collection of large quantities of data through the use of electronic health records, the discipline has seen a disproportionate amount of development in recent years (EHRs). An urgent need for knowledge production has arisen as a consequence of the vast amount of data that is now available, and in order to make use of this data, there is an urgent need for knowledge creation to take place. A significant role is played in this field of biosciences research by the application of data mining techniques and machine learning techniques, both of which are widely employed in the biosciences field, as well as by the employment of other methodologies. Chest pain is a symptom of CHD (Chronic Kidney

ISSN- 2394-5125 VOL 5, ISSUE 06, 2018

Disease), which is a medical condition in which the kidneys have been damaged and are no longer capable of filtering the blood in the manner in which they should be. Despite this, there are other sorts of renal illness that can occur, with this being the most common of them. Chronic kidney disease is caused by a combination of risk factors, including high blood pressure, type 2 diabetes, and a family history of renal sickness or failure. Other risk factors, such as obesity and smoking, can worsen the condition. Due to the fact that it is a long-term injury to the kidney, there is a considerable chance of renal failure in this condition. One of the most common medical conditions linked with kidney failure is cardiovascular disease (CAD), which accounts for approximately one-third of all reported cases of renal failure. Among the most significant issues that might arise in this circumstance, in addition to anaemia and bone disease, high potassium and calcium levels are among the most concerning.

When chronic kidney disease (CKD) is detected and treated early, it may result in a greater improvement in overall quality of life than when the illness is discovered and treated later in life. Early detection of chronic kidney disease (CKD) is crucial for successful treatment, and an accurate prediction method is needed to make this diagnosis at an early point in its progression. Chronic kidney disease (CKD) is a kidney condition that damages the kidneys over a long period of time. Researchers have revealed that a wide range of machine learning algorithms may be employed in a number of settings for the prediction of chronic renal disease as a result of scientific investigation. So yet, there has been no mention of this study in the literature. In order to forecast chronic kidney disease, a range of data preparation, data transformation, and classification approaches are employed in this work, which includes (CKD). The book also contains the most successful CKD prediction model yet established, which may be found in the appendix. The findings of this study suggest that doctors may be able to make more accurate predictions regarding the course of chronic renal disease in patients who are at an earlier stage of the disease as a result of these findings.

3. Existing System

The investigation and comparison of the performance of classifiers that are now in use with the performance of classifiers that are already in use is another option that might be considered in order to increase classification accuracy. Those suffering from Chronic Kidney Condition (CKD) benefit from early identification since it allows them to obtain timely treatment and prevent their disease from progressing further. Early detection of illness and treatment of patients are key components of the success of the medical industry, which is in desperate need of these services at this time of year. Techniques such as the use of new classifiers and the evaluation of their performances in order to discover better solutions for the target function in future work are conceivable and should be pursued in order to discover better solutions for the target function in future work.

ISSN- 2394-5125 VOL 5, ISSUE 06, 2018

The following are characteristics that are considered to be unfavourable.
When compared to its former performance, it is less efficient.
When playing this game, it is not possible to obtain an accuracy score.
a failure to recognise a sickness or to determine what kind of illness it is

3.1The System That Is Currently Under Investigation

According to the findings of this study, a number of different classification methods were used to determine whether or not a person had chronic renal disease. Specifically, the first two classification systems are applied to the topic under consideration: Classifiers were trained on the data in this experiment, and the support vector machine and the KNN classifier were the ones that were employed. Before being utilised in the classification of chronic renal illness, each classifier was trained on a data set consisting of patient records before being used in the classification. The accuracy, precision, and a measure of classification accuracy known as the F measure of classification accuracy were used to evaluate the performance of the classification algorithm.

The number 3.4 denotes the presence of a favourable factor.
As a result of the enhancements, the general efficiency of the system has been enhanced as well.
It is possible to reach an incredibly high level of precision.
Predict when the illness will reveal itself with a fair degree of accuracy.

4. SYSTEM STUDY

4.1 FEASIBILITY STUDY

The creation of a business proposal at this phase, which contains a high-level idea for the project as well as some preliminary cost estimates for the project's implementation, is the primary objective. It is decided whether the project is feasible at this point, and its viability is assessed at that point as well. Phase 3: A feasibility study of the proposed system will be carried out during the system analysis phase, which will be completed during the system analysis phase, and will be completed during the system analysis phase. This is done in advance of installation in order to ensure that the intended system does not become a financial burden on the organization's resources once it has been implemented. Before beginning a feasibility study, it is necessary to have a clear understanding of the fundamental requirements for the system.
One of the most significant aspects to examine while doing a feasibility study is the project's economic viability. The technical feasibility of the suggested solution is another significant point to consider when evaluating a solution.

ISSN- 2394-5125 VOL 5, ISSUE 06, 2018

EASIBILITY IN TERMS OF TECHNICAL IMPLICATION • CONFIDENCE IN A SOCIAL ENVIRONMENT

An economic assessment has concluded that the project is viable, and the project has been given that designation.

Specifically, the researchers think that this research is being conducted in order to determine what economic impact the system will have on the individual business in issue. According to the amount of money available for investment in research and development of the system that the company has accessible in comparison to other firms, the firm's resources are comparably limited. The costs must be backed up by supporting documentation if possible. As a result, the system that was developed was also completed on time and within budget, which was made possible by the fact that the vast majority of the technologies used were freely available on the internet at the time of construction. The system that was developed was also completed within budget. All that was required was the purchase of the products that had been altered. As stated in 4.1.2, technologically based solutions have a high likelihood of being operational in the near future.

This investigation is being carried out in order to determine the technological viability of the system under consideration, or more precisely, the technical requirements of the system under examination. As a result of the evolution of any system, there should be no considerable load placed on present technical resources. A tremendous amount of strain will be placed on the already available technical resources as a result of this development. As a result, the client will be required to operate at particularly high levels of quality as a result of the situation in question. Keeping the requirements as basic as feasible is critical when developing a system, so that when the system is put into operation, only little or no adjustments are required to ensure that the criteria are met.

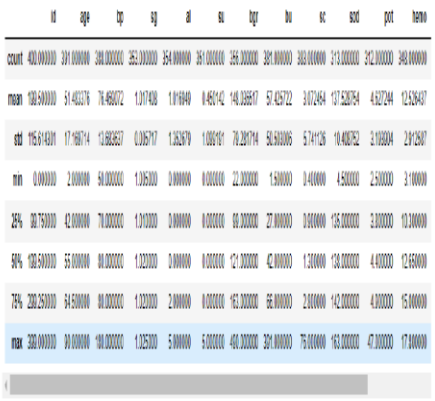
THE DETERMINATION OF SOCIAL FELIXIBILITY ASSESSING FEASIBILITY IN A SOCIAL ENVIRONMENT

It will be researched to what extent the system is acceptable to users from their point of view, and this will be one of the topics of inquiry that will be investigated. Instructional design is the process of training the user on how to make the most of the system's capabilities, such as increasing its performance to the maximum extent possible. According to the author, the system, on the other hand, should not be regarded as a risk by the user, but rather as something that must be tolerated as a necessary evil. There is no other aspect that can be taken into consideration when assessing whether or not a system will be accepted by its users. Given that he will be the final person to use the application, he will need to have increased self-confidence in order to be able to deliver a constructive evaluation, which will be much welcomed by everyone in his near area.

Result:

Data description:

ISSN- 2394-5125 VOL 5, ISSUE 06, 2018

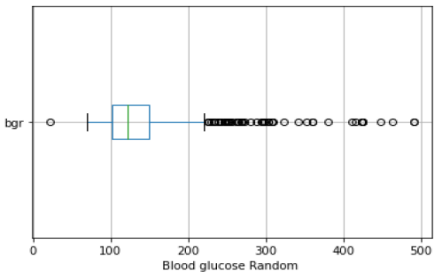


	id	age	bp	sg	al	su	bgr	bu	sc	sod	pot	hemo
count	400.000000	391.000000	388.000000	352.000000	354.000000	351.000000	345.000000	381.000000	383.000000	313.000000	312.000000	348.000000
mean	199.500000	51.483376	76.469072	1.017408	1.016949	0.450142	148.036517	57.425722	3.072454	137.528754	4.627244	12.526437
std	115.614301	17.169714	13.683637	0.005717	1.352679	1.099191	79.281714	50.503006	5.741126	10.408752	3.193904	2.912587
min	0.000000	2.000000	50.000000	1.005000	0.000000	0.000000	22.000000	1.500000	0.400000	4.500000	2.500000	3.100000
25%	99.750000	42.000000	70.000000	1.010000	0.000000	0.000000	99.000000	27.000000	0.900000	135.000000	3.800000	10.300000
50%	199.500000	55.000000	80.000000	1.020000	0.000000	0.000000	121.000000	42.000000	1.300000	138.000000	4.400000	12.650000
75%	299.250000	64.500000	90.000000	1.020000	2.000000	0.000000	163.000000	66.000000	2.800000	142.000000	4.900000	15.000000
max	399.000000	90.000000	180.000000	1.025000	5.000000	5.000000	490.000000	391.000000	76.000000	163.000000	47.000000	17.800000

Fig 10.6: Data description

pcv	wc	rc
330.000000	295.000000	270.000000
38.766667	8377.630508	4.700370
9.228118	2979.918951	1.029978
0.000000	1.000000	2.100000
32.000000	6500.000000	3.900000
40.000000	8000.000000	4.800000
45.000000	9800.000000	5.400000
54.000000	26400.000000	8.000000

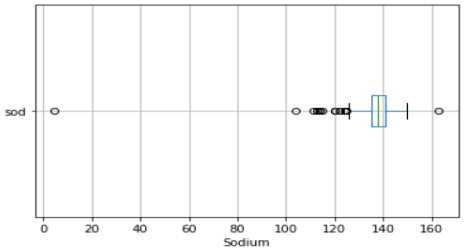
Blood Glucose Radom:



Blood Glucose Radom

Pottasium:

Sodium:



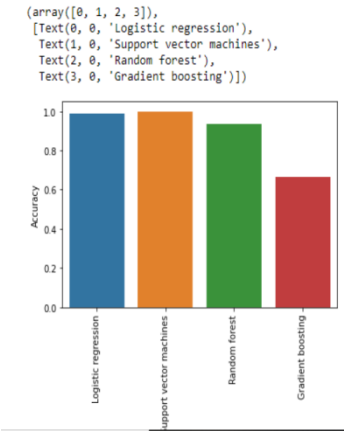
ISSN- 2394-5125 VOL 5, ISSUE 06, 2018

Sodium

Algorithms accuracy:

	Classifier	Accuracy	F1	Precision	Sensitivity	Specificity
0	Logistic regression	0.9875	98.75	98.79	98.11	1.0
1	Support vector machines	1.0000	52.80	43.89	1.00	0.0
2	Random forest	0.9375	1.00	1.00	1.00	1.0
3	Gradient boosting	0.6625	1.00	1.00	1.00	1.0

Algorithms accuracy



Accuracy result

5. CONCLUSION:

It is the goal of the researchers to determine whether it is feasible to detect chronic kidney disease (CKD) with the smallest amount of tests or characteristics possible by employing machine learning algorithms. In order to do this, a small dataset of 400 records is used, and the data is then categorised using four different machine learning classifiers, including logistic regression, SVM, random forest, and gradient boosting, among other techniques. A detailed analysis of the connections between variables has been carried out in order to decrease the number of characteristics and remove duplication where it is possible. A filter feature selection approach was applied to the remaining features and it was discovered that haemoglobin,

ISSN- 2394-5125 VOL 5, ISSUE 06, 2018

albumin, and specific gravity were the metrics having the greatest effect on predicting chronic kidney disease among the research participants.

REFERENCES

1. M. P. N. M. Wickramasinghe, D. M. Perera and K. A. D. C. P. Kahandawaarachchi, "Dietary prediction for patients with Chronic Kidney Disease (CKD) by considering blood potassium level using machine learning algorithms," 2017 IEEE Life Sciences Conference (LSC), Sydney, NSW, 2017, pp. 300-303.

2. 2. H. A. Wibawa, I. Malik and N. Bahtiar, "Evaluation of Kernel-Based Extreme Learning Machine Performance for Prediction of Chronic Kidney Disease," 2018 2nd International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, 2018, pp. 1-4

3. 3. U. N. Dulhare and M. Ayesha, "Extraction of action rules for chronic kidney disease using Naïve bayes classifier," 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Chennai, 2016, pp. 1-5.

4. 4. H. Zhang, C. Hung, W. C. Chu, P. Chiu and C. Y. Tang, "Chronic Kidney Disease Survival Prediction with Artificial Neural Networks," 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 2018, pp. 1351-1356

5. 5. J. Aljaaf et al., "Early Prediction of Chronic Kidney Disease Using Machine Learning Supported by Predictive Analytics," 2018 IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, 2018, pp. 1-9.

6. 6. Arif-Ul-Islam and S. H. Ripon, "Rule Induction and Prediction of Chronic Kidney Disease Using Boosting Classifiers, Ant-Miner and J48 Decision Tree," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, 2019, pp. 1-6.

7. G. Kaur and A. Sharma, "Predict chronic kidney disease using data mining algorithms in hadoop," 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, 2017, pp. 973-979.

8. N. Tazin, S. A. Sabab and M. T. Chowdhury, "Diagnosis of Chronic Kidney Disease using effective classification and feature selection technique," 2016 International Conference on Medical Engineering, Health Informatics and Technology (MediTec), Dhaka, 2016, pp. 1-6.