IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

MACHINE LEARNING BASED EARLY DETECTION OF KIDNEY DISEASE USING ECG SIGNALS

[1]Grandhi Lakshmi Prasanna, [2]Thyagarajan Prasad

[1]PG Student, [2]Head of the Department

[1]Digital Electronics and Communication Systems,

[1]Shree Institute of Technical Education, Tirupati, India

Abstract: This research article introduces the idea of detecting the presence of kidney disease through machine learning based classification modelling, by processing the patient's ECG signal. Recent studies and on-going researches have showed that patients undergoing kidney problems start developing cardiac problems-known as the Cardio Renal Syndrome (CRS) which can lead to a sudden cardiac arrest in the last stages of their disease. Since cardio-vascular diseases and the chronic kidney disease is inter-related, this model can be used for patients undergoing cardio-vascular problems to determine whether their kidneys have been effected or not. If the Chronic Kidney Disease (CKD) can be diagnosed at an earlier stage, it may give the patient some time to help reverse the disease or at least slow its progression by taking necessary medical steps. For this model, digitized ECG data was collected from open access databases such as PTB (kidney patients) and Fantasia (healthy people) from Physionet Database (www.physionet.org) and model was later validated using different data from the same online database.The validation process gave satisfactory results, as the model could successfully classify the users from being healthy or a kidney patient. In our study, we found an accuracy level of 97.6% which was the highest using both features QT and RR interval, in comparison to the accuracy that was found when either one of the features was used.

Index Terms **– Machine learning, Kidney disease, ECG signals, CRS, CKD**

I. INTRODUCTION

Chronic kidney disease includes conditions that damage your kidneys and decrease their ability to keep you healthy by doing the jobs. If kidney disease gets worse, wastes can build to high levels in your blood and make you feel sick. You may develop complications like high blood pressure, anemia (low blood count), weak bones, poor nutritional health and nerve damage. Also, kidney disease increases your risk of having heart and blood vessel disease. These problems may happen slowly over a long period of time. Chronic kidney disease may be caused by diabetes, high blood pressure and other disorders. Early detection and treatment can often keep chronic kidney disease from getting worse. When kidney disease progresses, it may eventually lead to kidney failure, which requires dialysis or a kidney transplant to maintain life.

Chronic kidney disease (CKD) refers to all 5 stages of kidney damage, from very mild damage in Stage 1 to complete kidney failure in Stage 5. The stages of kidney disease are based on how well the kidneys can do their job – to filter waste and extra fluid out of the blood. Diabetes and high blood pressure are the most common causes of CKD. If you have diabetes or high blood pressure, working with your doctor to keep your blood sugar and blood pressure under control is the best way to prevent kidney disease.

Recent studies and ongoing researches have showed that patients undergoing kidney problems start developing cardiac problems-scientifically known as the Cardio Renal Syndrome (CRS) which can lead to a sudden cardiac arrest in the last stages of their disease. Since cardio-vascular diseases and the chronic kidney disease is inter-related, this model can be used for patients undergoing cardiovascular problems to determine whether their kidneys have been effected or not. If the Chronic Kidney Disease (CKD) can be diagnosed at an earlier stage, it may give the patient some time to help reverse the disease or at least slow its progression by taking necessary medical steps.

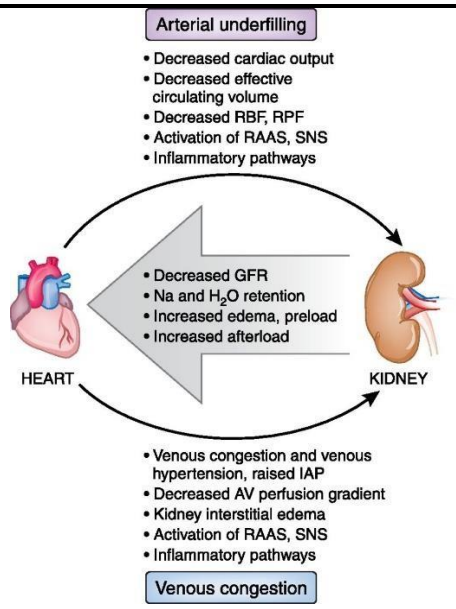


Figure 1: Cardio Renal Syndrome

II. LITERATURE SURVEY

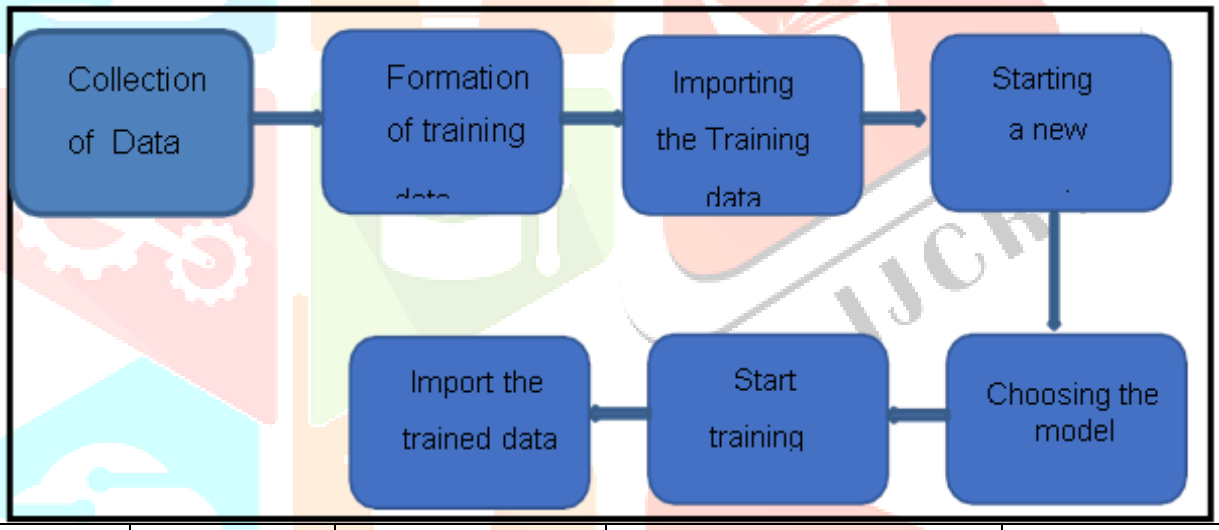
Author	Publication Year	Type of Kidney Disease	Techniques	Accuracy
Kaladhar	2012	Kidney stone	Naive Bayes	0.99%
			Logistic	1.00%
			J48	0.97%
			Random Forest	0.98%
			Naive Bayes	0.79%
			K-NN	0.7377 %
			Classification tree	0.9352 %
			C4.5	0.9352 %
			SVM	0.9198 %
			Random Forest	0.9352 %
K.R.Lakshmi	2014	Kidney dialysis	ANN	93.852 %
			Decision Tree(C5)	78.4455%
			Logical Regression	74.7438%
J.Van Eyck	2012	AK I	Gaussian process(aROC)	0.758%
			Gussian process(RMSER)	0.408%
Morteza Khavanin Zadeh	2012	Early AVF Failure	W-Simple Cart	85.11%
			WJ48	80.85%

Fig 2. Basic workflow of the model

To form the model, the first step was to extract digitalized ECG data from database. The digitized ECG was collected from two databases - the PTB database for the kidney patients‟ ECG and the Fantasia database for the healthy elderly patient's ECG as the CKD patients taken form the PTB database were all elderly, to reduce the ageing effect on CVD. The ECG signals were then processed using the Berger's algorithm, to find the required features- the QT interval and the RR interval.

III. IMPLEMENTATION

The following figure describes the basic work flowof the work done for this study.



er Y. Al- Hyari	2012	Chronic Kidney disease	Decision tree	--
Xudong Song	2012	Renal failure Hemodialysis	Decision tree	60-80%
N. Sri raam	2005	Kidney Dialysis	Association Rule	97.7%
Divya Jain	2014	Nephrotic syndrome(total protein)	C4.5	11% (error rate)
Jicksy Susan Jose	2012	Kidney Image	Association Rule	92%
			Navie Bayes	
koushal Kumar	2012	Kidney Stone	MLP	0.9613%
			LVQ	0.8459%
			RBF	0.8732

Using the extracted feature from the digitalized ECG, a training set was build where the patients were already labelled as ‘kidney’ or ‘healthy’. The labeled data is stored in classification learner app. Then the support vector machine is used for the classification of kidney disease.

The QRS complex is made up of three waves. These waves indicate the changing direction of the electrical stimulus as it passe s through the heart's conduction system. The largest wave in the QRS complex is the R wave.

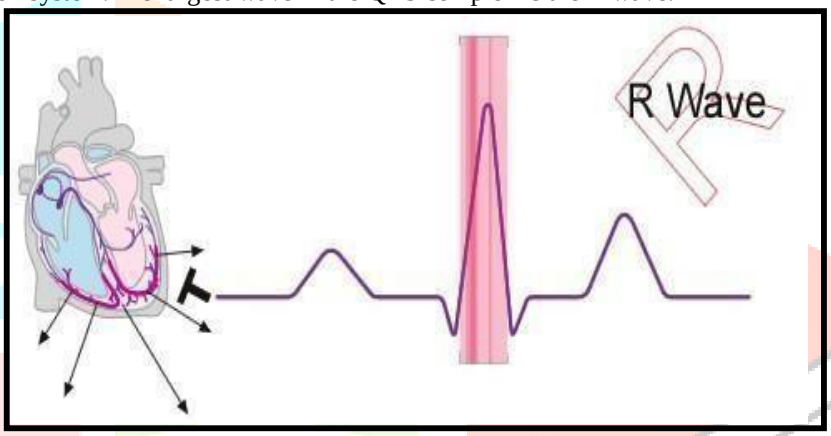


Fig 4: R wave

As you can see from the diagram, the R wave represents the electrical stimulus as it passes through the main portion of the ventricular walls. The wall of the ventricles are very thick due to the amount of work they have to do and, consequently, more voltage is required. This is why the R wave is by far the biggest wave generated during normal conduction.

Both ventricles repolarize before the cycle repeats itself and therefore a 3rd wave (t wave) is visible representing ventricu lar repolarization.

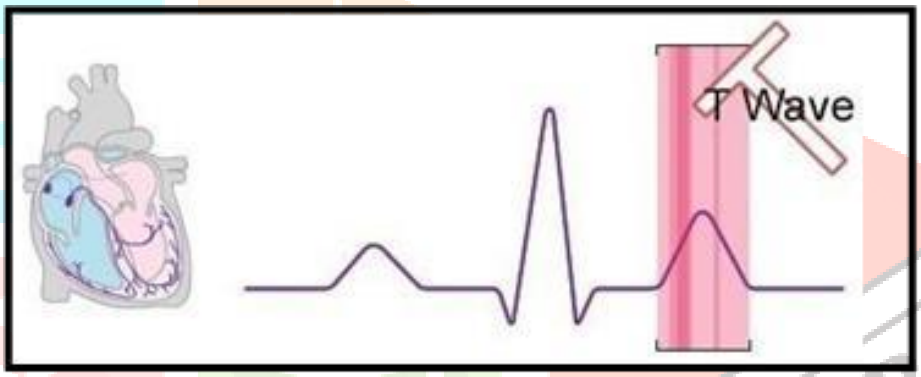


Fig 5: T wave

QT and RR interval:

This algorithm contains many leaves to make many fine distinctions between classes. This tree predicts classifications based on two predictors, x1 and x2. To predict, start at the top node. At each decision, check the values of the predictors to decide which branch to follow. When the branches reach a leaf node, the data is classified either as type 0 or 1.

TExpanded

1392x3 table

	1 QTinterval	2 RRinterval	3 Decision	4	5	6
1	0.2830	0.7040	kidney			
2	0.2890	0.6970	kidney			
3	0.3070	0.6980	kidney			
4	0.2880	0.7100	kidney			
5	0.3040	0.7050	kidney			
6	0.2880	0.7070	kidney			
7	0.3040	0.7090	kidney			
8	0.2860	0.7020	kidney			
9	0.2900	0.6990	kidney			

Fig 3.: training datasheet in the MATLAB workspace

The R Wave:

The T Wave:

The RR interval can be measured by calculating the distance between the first peak of R wave and the next peak of R wave in beats per minute. The QT interval can be measured by calculating the distance between the start of the QRS complex and the end of the T wave. Divide the QT interval by the square root of the RR interval.

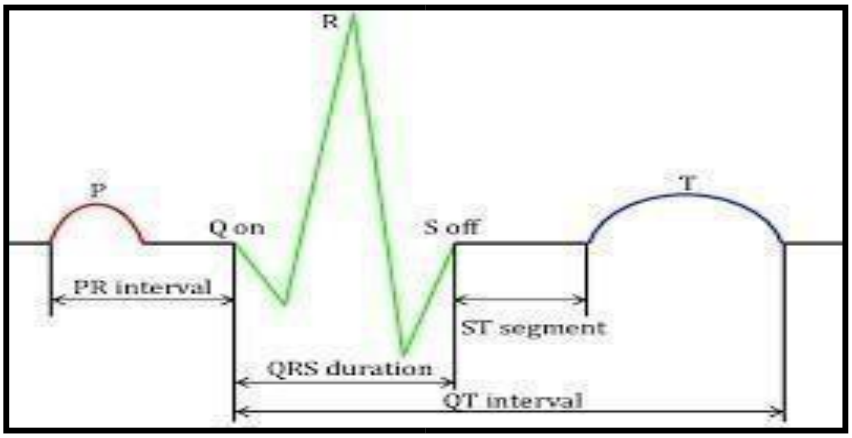


Fig 6: graph representing QT and RR intervals

Classification of disease:

Hence by extracting the QT and RR interval from the ECG signals by following the procedure mentioned above, these features are trained and tested by using support vector machine algorithm in the classification learner app.

Classification learner app:

The Classification Learner app trains models to classify data. Using this app, you can explore supervised machine learning using various classifiers. You can explore your data, select features, specify validation schemes, train models, and assess results. You can perform automated training to search for the best classification model type, including decision trees, discriminant analysis, support vector machines, logistic regression, nearest neighbors, naive Bayes, and ensemble classification.

On the Apps tab, in the Machine Learning group, click Classification Learner. Click New Session and select data from the workspace or from file. Specify a response variable and variables to use as predictors. On the Classification Learner tab, in the Model Type section, click All Quick-To-Train. This option will train all the model presets available for your data set that are fast to fit.

SVM (Support vector machine):

A support vector machine (SVM) is a supervised machine learning model that uses classification algorithms for two-group classification problems. The basics of Support Vector Machines and how it works are best understood with a simple example. Let's imagine we have two tags: red and blue, and our data has two features: x and y. We want a classifier that, given a pair of (x, y) coordinates, outputs if it's either red or blue. We plot our already labeled training data on a plane

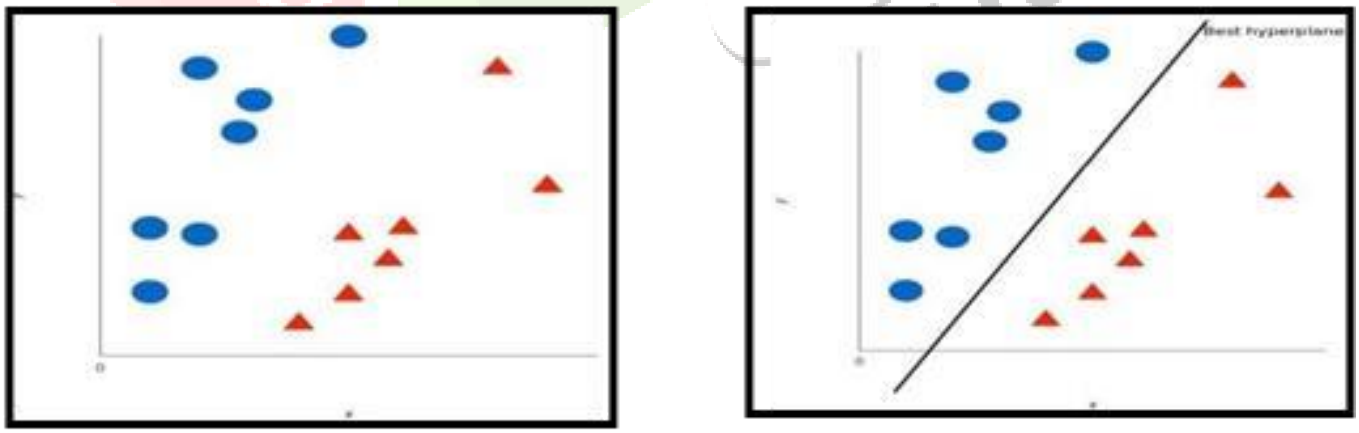


Fig 7. Test data without hyper plane

A support vector machine takes these data points and outputs the hyper plane (which in two dimensions it's simply a line) that best separates the tags. This line is the decision boundary: anything that falls to one side of it we will classify as blue, and anything that falls to the other as red. Then based on this hyper plane the test data is classified into one of the class i.e kidney or healthy

Hyper plane and Support Vectors in the SVM algorithm:

Hyper plane: There can be multiple lines/decision boundaries to segregate the classes in n- dimensional space, but we need to find out the best decision boundary that helps to classify the data points. This best boundary is known as the hyper plane of SVM. The dimensions of the hyper plane depend on the features present in the dataset, which means if there are 2 features (as shown in image), then hyper plane will be a straight line. And if there are 3 features, then hyper plane will be a 2-dimension plane. We always create a hyper plane that has a maximum margin, which means the maximum distance between the data points.

Support Vectors:

The data points or vectors that are the closest to the hyper plane and which affect the position of the hyper plane are termed as Support Vector. Since these vectors support the hyper plane, hence called a Support vector.

Confusion matrix:

Here the green boxes are the true values and then pink boxes are the false values. The upper green box represents the true positives i.e. the correctly classified healthy users and the lower green values are the true negatives which are the correctly classified kidney patients. The upper pink box represents the false negatives which are healthy people misclassified as kidney patients and the lower pink box represents the false positives which are kidney patients misclassified as healthy people. The three very important parameters to judge the performance of any algorithm are the sensitivity, selectivity and the accuracy and they can be all calculated easily from the plot

Out of all the different SVM models we have Fine Guassian SVM gives more accurate outcomes when compared to others.Hence we choose Fine Guassian SVM.

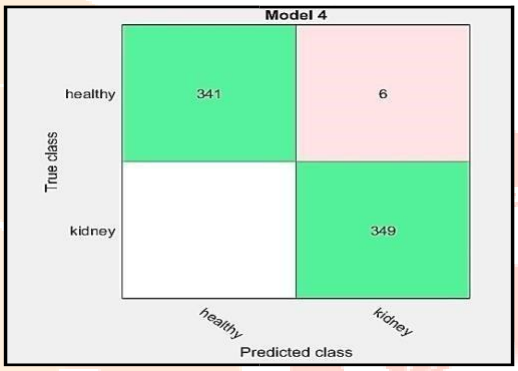


Fig 8.: confusion matrix of fine guassian SVM

IV. CONCLUSION

The research article can be used detect the presence of kidney disease through machine learning based classification modeling, by processing the patients ECG signal. Where the ECG signals are taken from open access databases such as PTB and fantasia from Physionet database. The SVM (Support vector machine) algorithm was used in order to find out whether the person is healthy or kidney patient. The experimental results are more accurate than the other traditional methods.

REFERENCES

[1] "About Chronic Kidney Disease", National Kidney Foundation, 2018. [Online]. Available: https://www.kidney.org/atoz/content/about-chronickidney disease. [Accessed: 26- Sep- 2018].

[2] B. Franczyk-Skóra, A. Gluba, M. Banach, D. Kozłowski, J. Małyszko and J. Rysz, "Prevention of sudden cardiac death in patients with chronic kidney disease", BMC Nephrology, vol. 13, no. 1, 2012

[3] "Global Facts: About Kidney Disease", the National Kidney Foundation, 2018. [Online]. Available: https://www.kidney.org/kidneydisease/global-facts-about-kidney-disease. [Accessed: 08- Apr- 2018].

[4]"Kidney Disease in Bangladesh", World Life Expectancy, 2018. [Online] Available: http://www.worldlifeexpectancy.com/bangladeshkidney-disease. [Accessed: 08- Apr- 2018].

[5]National Institutes of Health. "National Institute of Diabetes and Digestive and Kidney disease". Annual Data Report. Retrieved 22 November 2013.

[6] "Sudden Cardiac Death (Sudden Cardiac Arrest) | Cleveland Clinic", Cleveland Clinic, 2018. [Online]. Available: https://my.clevelandclinic.org/health/diseases/17522-sudden-cardiac deathsudden-cardiac- arrest. [Accessed: 12- Apr- 2018].

[7] Ronco, C.; McCullough, S.D. (2010). "Cardio-renal syndromes: Reports from the consensus conference of the acute dialysis quality initiative". European Heart Journal. 31 (6): 703–711. doi:10.1093/eurheartj/ehp507. PMC 2838681 Freely accessible. PMID 20037146

[8] Science Learning Hub. (2018). Label the heart. [Online] Available at: https://www.sciencelearn.org.nz/labelling_interactives/1label-the-heart [Accessed 26 Sep. 2018].

[9] "Sudden Cardiac Death (Sudden Cardiac Arrest) | Cleveland Clinic", Cleveland Clinic, 2018. [Online]. Available: https://my.clevelandclinic.org/health/diseases/17522-sudden- cardiac- deathsudden- cardiac-arrest. [Accessed: 12- Apr-

2018]. **[10]** D. Zachariah, P. Kalra and P. Roberts, "Sudden cardiac death in end stage renal disease: unlocking the mystery", Journal of Nephrology, vol. 28, no. 2, pp. 133-141, 2014.

[11] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/content/101/23/e215.full]; 2000 (June 13). **[12]** SF. Schoonjans, "ROC curve analysis with MedCalc", MedCalc, 2018. [Online]. Available: https://www.medcalc.org/manual/roc-curves.php. [Accessed: 09- Aug- 2018].

[13] R. D. Berger, E. K. Kasper, K. L. Baughman, E. Marban, H. Calkins, and G. F. Tomaselli, "Beat- to- beat QT interval variability: Novel evidence for repolarization liability in ischemic and non- ischemic dilated cardiomyopathy," Circulation, vol. 96, pp. 1557–1565, 1997.

[14] Evans, Frank. "Cardio-Renal Connections in Heart Failure and Cardiovascular Disease". NHLBI Working Group. Retrieved 22 November 2013.

[15] M. Corzo-Cuesta and C. Alvarado-Serrano, "An Algorithm for QT Dispersion Analysis: Validation and Application in Chronic Kidney Disease", 2016.