International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163
Volume 7, Issue 03 (March 2020) www.ijirae.com
Special Issue – 6th International National Level Conference- "MEEMIC-2020"

ECG SIGNAL BASED KIDNEY DISEASE PREDICTION SYSTEM USING MACHINE LEARNING

Prof.Muhammadu Sathik Raja,[1], Vinothini.G[2].
[1] Assistant Professor, Sengunthar College of Engineering, Tiruchengode.
PG Scholar[2]
Department of Medical Electronics Engineering,
Sengunthar College of Engineering. Tiruchengode,Tamil Nadu, INDIA
[1]Scewsadik@gmail.com, [2]gvkvinothini@gmail.com

Manuscript History
Number: **IJIRAE/RS/Vol.07/Issue03/Special Issue/24.MRAESCE10103**
Received: 15, February 2020
Final Correction: 27, February 2020
Final Accepted: 10, March 2020
Published: **14, March 2020**
Editor: Dr.A.Arul Lawrence selvakumar, Chief Editor, IJIRAE, AM Publications, India

Abstract: This paper 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 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 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 (for kidney patients) and Fantasia (for healthy people) from Physionet Database (www.physionet.org) and the model was later validated using different data from the same online database. 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.
Keywords: chronic kidney disease (CKD), cardio-vascular problems, cardio renal syndrome (CRS), sudden cardiac death (SCD), MATLAB

INTRODUCTION

Chronic Kidney Disease (CKD), or the chronic renal failure, is a disease where the kidneys start to lose their functionality [1]. It is a chronic condition which causes the kidney to deteriorate and lose their ability to function properly and lastly lead to the fifth and final, fatal stage- the End Stage Renal Disease (ESRD), where the kidney functions drop to almost 10 to 15 percent of their healthy capacities [2]. When the disease has progressed to this stage, a kidney transplant or dialysis are the only options for the patient to survive. Studies have since shown that, amongst the CKD patients' death, 60% of the deaths are Sudden Cardiac Deaths (SCD) whereas the rest 40% are other cardiovascular mortalities [2]. In accordance to a report from the US National Kidney Foundation, almost 10% of the world population suffers from CKD, among which around 2 million people require dialysis or a transplant to live [3]. Narrowing down more locally, a leading daily reported that in Bangladesh, one out every seven people suffer from kidney problems and 3.24% of the population death can be traced back to kidney disease [4]. Sudden Cardiac Death (SCD) is defined as the unexpected natural death due to a cardiac cause, in a person that does not have any prior potential fatal condition [6]. It can occur due to rhythm abnormalities in the heart, known as arrhythmias.

When the heart undergoes ventricular fibrillation, an arrhythmia where the heart fires erratic, chaotic impulses from the ventricles of the heart which interrupts the sinus rhythm and thus the normal blood flow of the heart.

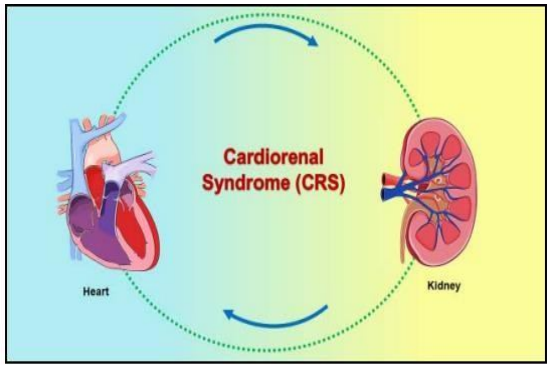

Fig. 1. Cardio Renal Syndrome (CRS)

This, in turn, causes a depletion of oxygen in all parts of the body. The ventricular fibrillation stage requires strict medical attention; otherwise the patient might succumb to SCD [6]. This phenomenon, of a kidney condition causing cardiac problems and vice versa is medically termed as the Cardio Renal Syndrome (CRS). Since the heart and kidneys are connected through various pathways that ensure a stable blood flow, a problem in one often induces a problem in the other [7]. Thus, in reference to the CRS, it is possible to say that patients undergoing symptoms of cardio vascular diseases (CVD) may also be suffering from CKD. Since it is already known that different CVDs leave characteristics traces in the patient's ECG, it is also possible to detect the presence of CKD from the same ECG of the patient if the traces for CKD are known. From various important studies, it was found that ECG of any patient undergoing CKD, shows some significant changes which can be traced back to CKD [9, 10]. The QT duration was seen prolonged in CKD patients such that 460ms in female while 450ms in men and the QRS amplitude was also seen to be increased by approximately 0.18Mv. The traditional method of detecting kidney diseases are the invasive methods of blood tests (e.g. GFR test) or kidney biopsy, and the non-invasive methods are ultra sound imaging or urine tests to check for creatinine levels, and End stage CKD patients. Thus, this model opens up an option for patients to detect their kidney disease through a simple non-invasive way by means of their available bio signals i.e. ECG signal. ECG signal is widely used for CVD analysis as almost all types the heart abnormalities can be detected from this. Since, any patient with CVD are very likely to undergo an ECG test, through this model, the presence of CKD can also be detected from the same ECG, by checking for prolonged QT intervals and also checking RR intervals which can be used to measure the HRV. The following sections of this paper will discuss about the ECG data processing, model formation and validation stages.

DATA INFORMATION

For the training of the model, all the data was collected from two open access online database, PTB (used for digitized ECG of the kidney patients) and Fantasia (used for the digitized ECG of healthy people) from Physionet database (www.physionet.org). The database each contained two minute long digitalized ECG signal, of the patients, from which the two required features- QT interval and the RR interval was extracted using Berger's algorithm [13]. The patients, whose ECG were taken had an average age between 50 to 70 years old [11, 12]. For the PTB database, 290 subjects were considered both male and female with a mean age of 57.2, 549 records were collected by taking 5 records for each subject. Each of these records included 15 simultaneously measured signals (the standard 12 leads, 3 Frank lead ECGs), which were digitized at a sampling rate of 1000 samples per second with 16-bit resolution. The sampling rate was also varied up to 10 KHz. From this database we choose 7 subjects who were diagnosed with both CVD and CKD as presented in [15]. . This was then digitized at a sampling frequency of 250Hz and using an automated arrhythmia detection algorithm, each heart beat was annotated and finally verified by visual observation [11 ,12]. The following table summarizes brief information of the seven patients whose times series ECG were used for the training of the model.

TABLE I. INFORMATION OF KIDNEY PATIENTS [15]

Patient Intex	Age	Cardiovascular Problem
Patient 12	71	Myocardial Infraction
Patient 13	67	Myiocardial Infration
Patient 19	50	Cardio Myopathy
Patient 70	76	Hypertropathy
Patient 85	35	Mynocarial Infraction
Patient 69	73	Myocardial Infraction

METHODOLOGY

The following figure describes the basic work flow of the work done for this paper.

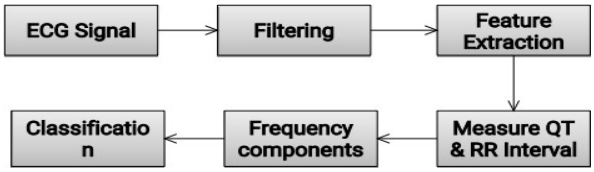

Fig. 2. Basic work flow of the model

To form the model, the first step was to extract digitalized ECG data from database. As discussed in the previous section, digitized ECG was collected from two databases - the PTB database for the kidney patients' ECG and the Fantasia database for the healthy elderly patients' ECG as the CKD patients taken from 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. Using the extracted feature from the digitalized ECG, a training set was build where the patients were already labelled as 'kidney' or 'healthy'.

Patient Intex	Age	Cardiovascular Problem
Patient 12	71	Myocardial Infraction
Patient 13	67	Myiocardial Infration
Patient 19	50	Cardio Myopathy
Patient 70	76	Hypertropathy
Patient 85	35	Mynocarial Infraction
Patient 69	73	Myocardial Infraction

The version of MATLAB used was 2014b since the previous versions might not have the above mentioned application. The created training set was then imported into the Classification Learner application, where the parameters for the model, i.e. the algorithm- linear Support Vector Machine (SVM), features used- QT interval and RR interval, the kernel function- linear, automatic and the cross-validation scheme was selected and the model was trained. It is very important to note that, the supervised machine learning was used because the training data set contained pre-labelled data and the ECG for both groups of patients were already known. Under supervised machine learning, SVM was chosen because it showed good performance in many studies for classification purposes. The trained model, once imported to the workspace was in the form of a function where a new table format data was passed, and the function was used to classify the new data. Fantasia were again taken and tabulated into another excel file where the features were there but the decision was missing.

RESULTS ANALYSIS

A summary of the model that was returned is that the accuracy of the model is 97.6% and it only took approximately 10-15 seconds to get trained. The validation for the model was done using data from the aforementioned online database and it was seen that the model could classify most of the patients correctly.

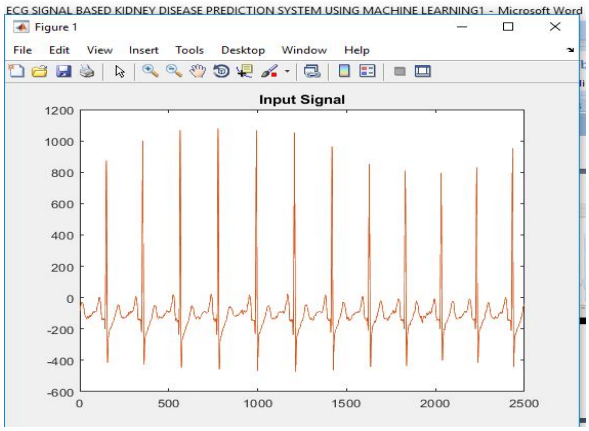

Fig. 3. An example of the training datasheet in the MATLAB workspace

The model can further be explained through the various plots that can describe its characteristics such as the scatter plot and the confusion matrix. The following figure is the scatter plot of the model; the red color is used to mark the points that represent kidney patients, while the blue color is the representatives of the healthy patients. i.e. the points that were supposed to be in one class got labeled as the other. Since the model has an accuracy of 97.6%, it can be seen that the majority of the points have been correctly classified.

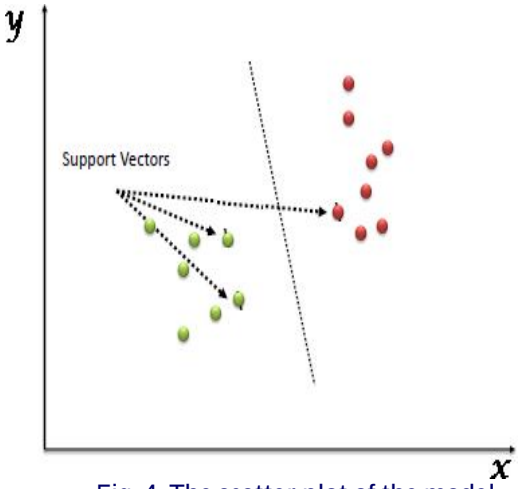

Fig. 4. The scatter plot of the model

The following figure is the confusion matrix for the model. 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.

CONCLUSION

Chronic Kidney Disease (CKD) is a worldwide epidemic in these days, which leads to millions of deaths each year. This is because CKD, when advances can lead to many CVD such as SCD in the last stages. In reference to CRS, many people suffering from cardio-vascular problems may also be a victim of the CKD, and sometimes the treatment for CVDs may also get limited due to the presence of CKDs [14]. Thus a classification based model has been developed, to be able to detect kidney disease in the early stages, from their digitalized ECG, using machine learning algorithms. It takes two concerning features- QT interval and RR interval to detect the presence of CKD in the patients, with an accuracy of 97.6%. The primary advantage of this model is the fact that it provides a safe non-invasive way for patients to determine the state of their kidneys. Since, all types of CVD can be diagnosed from the ECG of the patient and any patient suffering from any sorts of CVD, must undergo an ECG test, the same test signal can be passed onto the model to determine if the patient's kidneys are getting affected or not

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. Raja, G P& Mangai, S 2018, 'Investigation On Optimization, Prioritizing and Weight Allocation Techniques for Load Balancing and Controlling Multimedia Traffic in Wireless Mesh Network', International Journal of Business Information Systems, SCOPUS Indexed Journal (Inderscience) - (P ISSN No: 1746-0972). Published Online: 10th Feb 2020, DOI: 10.1504/IJBIS.2020.105161.IF: 0.72.
6. Geetha. E & Nagarajan. C , 2015, 'Embedded controller integrated into a wireless sensor network for induction motor parameter monitoring using ZigBee, international Journal of Applied Engineering Research, Special Issue, Volume 10, No.9 2015, pp.9343-9346.

7. Dr.R.Satish Kumar and Dr.K.Umadevi " A Novel peak torque Excitation Technique for Torque Improvement in Exterior Rotor Permanent Magnet Brushless DC Motor", International journal of Innovative research in Advanced Engineering, Vol. 1, 2014, pp. 227-236 (Impact factor 1.311).
8. Dr.R.Satish Kumar and Dr.K.Umadevi " Torque Improvement for an Exterior Rotor PermanentMagnet Brushless DC Motor", International journal of Innovative research in Advanced Engineering, Vol. 1, 2014, pp. 1-5 (Impact factor 1.311).
9. Dr.R.Satish Kumar and Dr.K.Umadevi " Novel Technique for Measurements of Dielectric Properties and Microwave Heating of In-Shell Eggs without Explosions in Microwave Oven for Pasteurization", International journal of Innovative research in Advanced Engineering, Vol. 2, , 2015, pp. 69-77 (Impact factor 1.311).
10. Dr.R.Satish Kumar and Dr.M. Y. Sanavullah" Theoretical and experimental study of cooking regions for shell eggs in a domestic Microwave oven", International Conference on Electronics Computer Technology, 2011, https://doi:10.1109/ICECTECH.2011.5941909

