International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 08 | Aug 2020 www.irjet.net p-ISSN: 2395-0072

A GENERIC REAL TIME APPLICATION FOR CKD PREDICTION USING MACHINE LEARNING

Dr. Nagaraju C[1], Varun.B[2]

[1]Assistant Professor, [2]Student
[1,2]Department of Electronics and Communication, The National institute of Engineering (NIE), Mysore, Karnataka, India.

Abstract - *Chronic kidney disease (CKD), is the gradual and irrevocable destruction of kidneys. It reduces the potential of humans to stay healthy. The diagnosis of CKD Start with a medical history. Discovery and analysis ought to be done earlier so it will ordinarily shield renal disorder from acquiring a worse condition. Here authors have utilized Machine Learning Techniques into actions for prediction of CKD. Authors examine the presentation of Naive Bayes, K-Nearest Neighbors (KNN) and GFR strategy for stage prediction based on its exactness, accuracy and execution time for CKD forecast. Comparatively Naïve Bayes shows good results than KNN and stage prediction is done at every instance when CKD is predicted.*

Key Words: **Chronic, Diagnosis, kidney disease, Machine Learning, Stage prediction**

1.INTRODUCTION

AI (ML) is a type of Artificial insight that engages a framework to gain from information instead of through unequivocal programming. As the calculations absorb preparing information. It's conceivable to create exact models dependent on the informational index. A ML model is the yield created when we train the AI calculation with information.

ML procedures are required to improve the exactness of prescient models. Contingent upon the idea of issue ML is grouped into various classes.

1. Supervised learning – it starts with an all-around Defined arrangement of information and how it has been characterized. This classifier is proposed to discover designs in information that can be applied to an investigation procedure.

2. Unsupervised learning – it is utilized when the issue requires a huge measure of unlabelled information. It directs an iterative procedure, breaking down information without human association.

3. Reinforcement learning – it is inclining model. Here the framework isn't prepared with the example information.

4. Deep leaning – this technique joins neural systems in progressive layers to gain from information in an iterative way.

Huge information in setting ML requires right arrangement of information to apply to a learning procedure. Large information can assist with improving the precision of ML models.

Different information mining characterization approaches and ML calculations are applied for forecast of unending sickness. Here we are worried about CKD. It is where kidneys become harmed and can't channel poisonous in the body

2. LITERATURE SURVEY

Anusorn Charleonnan, Thipwan Fufaung, TippawanNiyomwong,WandeeChokchueypatt anakit,[1] Sathit Suwannawach, Nitat Ninchawee have performed a research in 2016 on, "Predictive Analytics for Chronic Kidney Disease Using Machine Learning Techniques". The prescient models by utilizing ML strategies including K-nearest neighbours (KNN), support vector machine (SVM), logistic regression (LR), and decision tree classifiers to anticipate CKD. From the trial results, it tends to be seen that SVM classifier gives the most elevated precision. What's more, SVM has most elevated affectability in the wake of preparing and testing by the proposed technique. Her creators have inferred that SVM classifier is appropriated for foreseeing the incessant kidney ailment.

Helmie Arif, Indra Malik and Nurdin Bahtiar [2] have performed a research in 2018 to "Evaluate the

Kernel-Based Extreme Learning Machine Performance for Prediction of Chronic Kidney Disease". In this paper creators have proposed a model for anticipating Chronical Kidney Disease, including standard ELM and Kernel-Based ELM. Five strategies were assessed in six situation including standard ELM, Linear ELM, Polynomial-ELM, RBF-ELM, and Wavelet-ELM. As indicated by them RBFELM utilizing all the highlights and utilizing chosen highlights, gives the best execution in anticipating Chronic Kidney Disease. The subsequent affectability and explicitness arrived at 99.38% and 100% individually.

Devika R,Sai Vaishnavi Avilala, and V.Subramaniyaswamy [3] have presented a research in 2019, on Comparative Study of Classifier for Chronic Kidney Disease prediction using Naive Bayes, KNN and Random Forest. In this paper, constant kidney infirmity is foreseen utilizing explicit classifiers and a relative investigation of their general execution has been actualized. From the assessment, creators found that, out of classifiers Naive Bayes, Random Forest, and KNN, Random Forest classifier performed higher than the other option. The cost of forecast of CKD is progressed.

Lambodar Jena and Ramakrushna Swain [4] have performed a research in 2017, on Chronic Disease Risk Prediction using Distributed Machine Learning Classifiers. The principle goal of this paper is to anticipate interminable kidney ailment. Here creators have utilized two calculations for example Naive Bayes and Multilayer Perceptron for the experimentation reason. These calculations are actualized utilizing WEKA AI instrument to break down. Exactness which is gotten in the wake of running these calculations in the yield window. In the wake of running these calculations, the yields are looked at based on precision accomplished. These calculations have been contrasted with characterization exactness with one another based on effectively grouped cases, mean total blunder, Kappa measurements and RMSE metric. The outcomes show that MLP classifier outflanks Naïve Bayes classifier in all part as for the boundaries determined. It is inferred that MLP classifier is the best expectation calculation for constant kidney sickness conclusion.

Ahmed J. Aljaaf,Dhiya Al-Jumeily,Hussein M. Haglan,Mohamed Alloghani, Thar Baker,Abir J. Hussain and Jamila Mustana [5] have presented a research in 2014, on Early Prediction of Chronic Kidney Disease Using Machine Learning Supported by Predictive Analytics. In this test methodology, they have applied two discovery models for early expectation of CKD, for example SVM and MLP model MLP neural system design shows the most noteworthy AUC of 0.995 and TPR of 0.9897.

Gunarathne, Perera, and ahandawaarachchi [6] have presented a research in 2017 on Performance Evaluation on Machine Learning Classification Techniques for Disease Classification and Forecasting through Data Analytics for Chronic Kidney Disease (CKD). Creators have seen that the Multiclass Decision Forest calculation gives the most noteworthy precision of 99.1%. As indicated by their examination on results, utilizing the 14 properties of CKD patients it can anticipate the CKD status of new patients with a 99.1% exactness utilizing their model. Creators have mostly centred around to actualize a model to distinguish another CKD patient's wellbeing condition by centering progressively included regions which will assist with having a superior thought regarding patient's condition.

AKM Shahariar Azad Rabby,Rezwana Mamata, MoniraAkter Laboni, Ohidujjaman and Sheikh Abujar [7] have presented a research in 2019, on Machine Learning applied to Kidney Disease Prediction: Comparison Study. Creators point is to develop an updated and unmistakable AI (ML) application that can feasibly see and foresee the condition of a wearisome kidney disease. In this work, the ten most fundamental AI game plan systems were considered by the creators for predicting never-ending kidney infection. In the wake of finishing this, we can include more classifications in this work, can make this progressively proficient. Utilizing more classifiers on this informational index can show signs of improvement understanding on which classifier.

By surveying all these papers, we have overcome certain limitations. In all these papers the dataset used is minimal like around 400 instances with limited parameters. In our proposed work we have used around 1500 instances with 24 different parameters taking every record into consideration in order to get the accuracy in well-defined way. Here we have also using GFR technique for stage prediction. More the number of Dataset, more the accuracy in predicting the disease.

Our work dominatingly centers around identifying hazardous sicknesses like CKD utilizing directed learning algorithms. Hence, the forecast of constant kidney infection is one of the most significant errands. Incessant Kidney sickness is anticipated utilizing grouping strategies for information mining. The classifiers utilized here are, Naive Bayes, K-Nearest Neighbors (KNN) and GFR procedure for stage expectation. Their exhibition is surveyed as far as exactness and accuracy.

3. PROPOSED METHODOLGY

In our proposed framework two Machine learning methods are utilized to foresee the nearness of ceaseless kidney malady (CKD) in people. Naive Bayes, k-Nearest Neighbors (KNN), and Glomerular filtration rate (GFR) procedure for stage expectation are the significant techniques examined in the proposed strategy. The data set for constant ailment was assembled and applied on each classifier to foresee the sickness and along these lines the exhibition of the classifier is inspected dependent on their exactness and accuracy. The plan activity is as per the following: The dataset for CKD patients are gathered and taken care of into the classifier named Naive Bayes, KNN and GFR strategy. The expectation of CKD is finished with the calculations acted in C Sharp Language. In this paper, the informational collection is gathered from the UCI ML store with the dataset of 1500 examples, the contribution for forecast. The dataset comprises of qualities and qualities. as shown in the Table 1. Fig (1) Shows how exactly our proposed methodology works using ML Technique with the dataset that we have used in the proposed method.

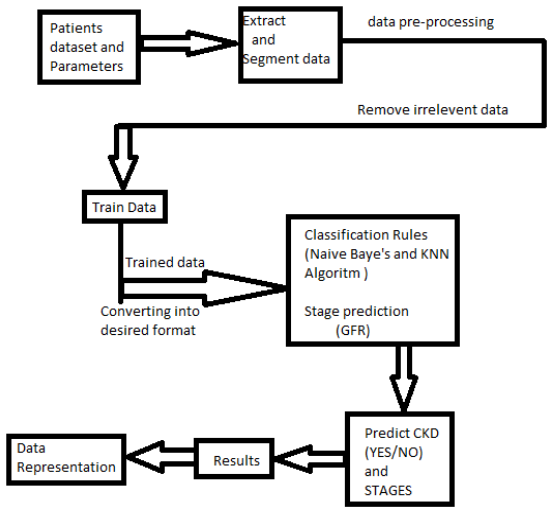

Fig1.Chronickidney disease prediction Process

Finally, from the result the accuracy of algorithms will be enhanced. The experimental results retrieved, that shows the most effective classifier among the methods.

4. MATERIALS AND METHODS

A. Data collection and Attributes

In Machine learning the quantity and quality of our data shows how accurate our model is. Outcome is generally a representation data which we will use for training. In order to empower the software to work independently and generate solutions it's the duty of researcher to perform the following action like finding required algorithms and required data must be collected and fed into the system in advance.

The heart of the machine learning process is data collection. And for our paper, there was no alternative of data. So, it has become our most challenging task for our proposed method.

Here we have used publicly available dataset which is downloaded from the UCI repository. This dataset includes 1500 patient records. The CKD data set consists of 24 parameters (i.e. predictors) in addition to the binary class attribute.

As illustrated in table 1. 24 different parameters have been listed which includes attributes like Age, blood

pressure etc. and are specific with the values used. These 24 parameters are compared with the new patient dataset and prediction is done.

Table 1. Attributes and values used

Attributes	Values used
Age	Discrete whole number qualities
Blood pressure	Discrete whole number qualities
Specific gravity	Numeric qualities
Albumin	maturity Values (0,1,2,3,4,5)
Sugar	maturity Values (0,1,2,3,4,5)
Red blood	maturity Values (Normal, abnormal)
Pus cell	maturity Values (Normal, abnormal)
Pus cell clumps	maturity Values (present not present)
Bacteria	maturity Values (present not present)
Blood glucose random	Numeric qualities
Blood urea	Numeric qualities
Serum creatine	Numeric qualities
Sodium	Numeric qualities
Potassium	Numeric qualities
Hemoglobin	Numeric qualities
Packed cell volume	Numeric qualities
WBC count	Discrete whole number qualities
Hypertension	maturity values (yes, no)
Diabetes	maturity values (yes, no)
Coronary artery disease	maturity values (yes, no)
Appetite	Maturity values (good, poor)
Pedal edema	maturity values (yes, no)
Anemia	Nominal values (yes, no)

B. Analysis

A thorough analysis of the data is performed in this section in order to explain its essence or to determine its important characteristics. This is a major point towards a compelling and legitimate expectation of CKD. The structure of ML process is helped out through information assortment and examination which is as appeared in the fig (2). Regularly, the more grounded the importance of a boundary to the class quality implies that this boundary is fundamental for an ideal learning execution and predication. Along these lines, it is beneficial to dissect input boundaries to define their unfair force in the expectation of CKD in the beginning phase. This progression empowers us to comprehend the degree of cover among CKD and solid people as for certain boundary.

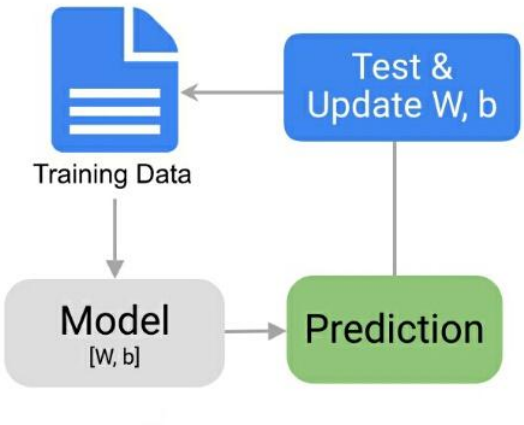

Fig.2. Frameworks for approaching the machine learning process

Proposed method includes hard and software requirements towards the prediction of the disease using machine learning approach.

1. Hardware Requirements:

Hardware required to develop the module is as listed below

Processor : Pentium IV onwards

RAM : 2GB +

Hard disk space : 40GB +

Standard PC configuration to carryout challenging computing

2. Software Requirements:

Software required in development is as listed below

Operating System : Windows XP version, 7 & Higher

Framework : DOTNET

Design Tool : Visual Studio 2010.

Language : C#.NET

Back End : SQL Severest CKD.

5. ALGORITHMS

A. Naïve Bayes algorithm

Stage 1: Scan the dataset (stockpiling servers) to retrieve the necessary mining information from the servers, such as database, cloud, exceed expectations sheet and so on.

Stage 2: Calculate the likelihood of each trait esteem [n, n_c, m, p].

Here for each trait we ascertain the likelihood of event utilizing the accompanying equation. For each class (infection) we ought to apply the formulae.

Stage 3: Probability to distinguish CKD

P (trait esteem (ai)/subject worth vj) = (n_c + mp)/(n+m)

Where:

n = the quantity of preparing models for which v = vj

nc = number of models for which v = vj and a = ai

P= from an earlier P(aijvj) gauge

m = the comparable example size

Stage 4: Multiply the probabilities (P) by p

for each class, here we increase the after effects of each property with p and conclusive outcomes are utilized for characterization.

Stage 5: Compare the qualities and arrange the credit esteems to one of the predefined sets of class. In this method probability of CKD is found using equation (1).

P= [n_c + (m*p)]/ (n+m) ----------(1)

B. k-Nearest Neighbors (KNN) Algorithm:

Stage 1: Determine K (no of closest neighbors).

Stage 2: Calculate separation (Euclidian).

Stage 3: Determine K-least separation neighbors.

Stage 4: Gather class Y estimations of closest neighbors.

Stage 5: Use straightforward dominant part of closest neighbors to anticipate estimation of question case. Its formula is classified as below:

Euclidean = $\sqrt{\sum} (\boldsymbol{xi} - \boldsymbol{yi})\mathbf{2}\ \boldsymbol{k}\ \boldsymbol{i}\mathbf{=1}$ ----------(2)

C.GFR Technique for Stage Prediction:

Here GFR is mainly used to find the level of CKD through different stages.

A*(Scr/B) ^ C * (0.933) ^ age ---------(3)

A, B, C are Constants.

In the proposed method GFR will be able to detect five different stages of kidney disease and they are as listed below.

phase 1-Popular or Strong (GFR > 90 mL / min)

Phase 2-Mild CKD (60-89 mL / min GFR)

The 3A-CKD stage (GFR = 45-59 mL / min)

3B-Moderate stage CKD (GFR=30-44 mL / min.

Phase 4-Solid CKD (GFR = 15-29 mL / min)

Phase 5 - End Stage CKD (GFR)

6. RESULTS AND DISCUSSIONS

This area presents a point by point results and conversation acquired by the creators in the current work. The outcomes are introduced as tables and previews. To anticipate the CKD Naive Bayes Algorithm are K-nearest neighbors (KNN) utilized lastly GFR is utilized for stage forecast. The proposed technique is applied on in excess of 1500 informational indexes with various qualities. Investigating these three strategies we acquire the accompanying outcomes as demonstrated as follows.

The Naïve Bayes Analysis (NBA) with the constraints and its accurate results. Here all 1500 datasets have been tested on the basis of 24 parameters that we

have taken for the test. NBA got 91.54% accuracy on the test set. It has been shown graphically in fig. 3.

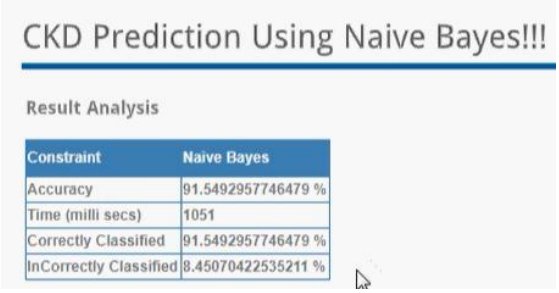

Fig. 3 CKD Prediction using Naïve Bayes

K-nearest neighbors (KNN) analysis with all constraints and results. Here also same dataset and parameters are used to find the results. From the test KNN got 87.32 accuracy on the test set and it has been recorded graphically as in fig. 4.

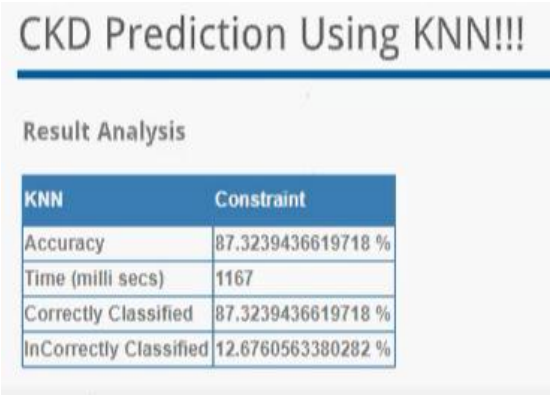

Fig. 4 CKD Prediction using KNN

GFR technique basically predicts 5 different stages of CKD in which we display some of the examples of it. A typical stage 2 and stage 3 outcomes are shown in fig (5) and fig. (6) respectively.

Fig. 5 Stage Prediction using GFR

This result shows that CKD patient with the specific stage based on which the patient has been diagnosed with CKD with the available attributes. As mentioned earlier after the patient is confirmed with CKD algorithm will predict the stage of the disease with 3 parameters (age, gender and serum creatine). These three parameters help in finding stages of the disease.

Fig. 6 Stage Prediction using GFR

Similarly, as mentioned in the GFR algorithm 5 different stages of CKD can be found using GFR

technique where each stage has its own specific GFR value as mentioned earlier.

In this way using GFR technique we can treat the patient as early as possible in right manner and cure the patient as soon as possible with the right techniques implemented by the doctors. This helps the doctor to take a quicker action on the disease at early stage.

After all these prediction techniques the treatment will be uploaded by the doctors in the portal which will be more helpful for both patient and the doctors in order to check the patient history. Patient can also view the treatment details. Hence, it's a medical sector application where it is useful for the mankind in a better way.

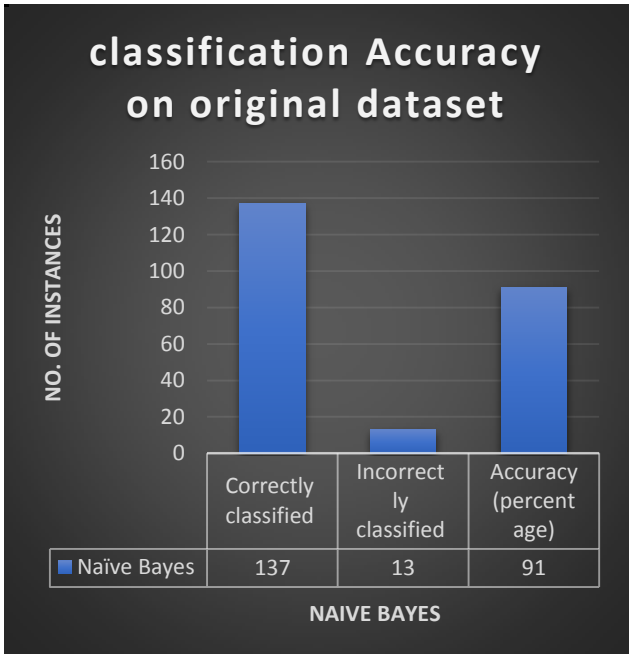

Fig.7 Graph showing Naïve Bayes classification Accuracy on original dataset.

In the proposed method we have undertaken an experimental procedure in order to predict the CKD by predictive analysis using machine learning techniques. Two class labels used as targets in the study (i.e. patients with CKD and patients with NOT CKD), over which two machine-learning methods were simulated using Naïve-Bayes Algorithm and K-nearest neighbors (KNN). Accuracy of Naïve-Bayes Algorithm was found as 91% and has been graphically shown in fig (7). Similarly, for K-nearest neighbors (KNN) it was found to be with 87% which has shown in fig (8). In comparison of these Naïve-Bayes Algorithm, out performs with K-nearest neighbors towards the detection of CKD.

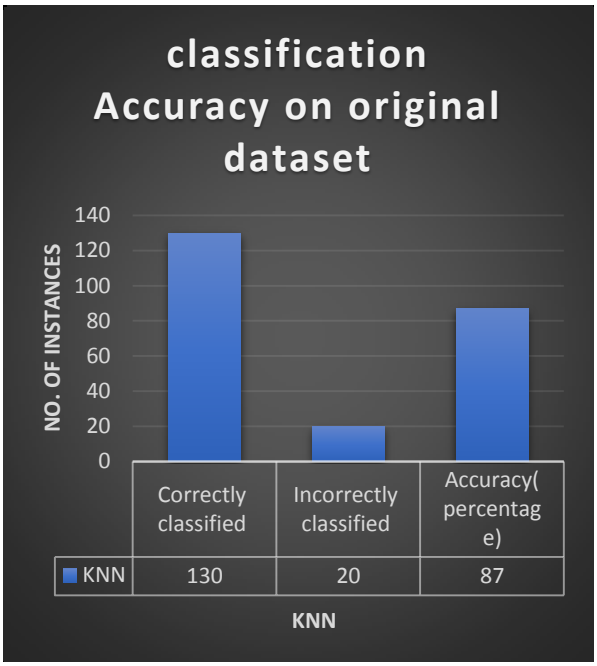

Fig.8 Graph showing KNN classification Accuracy on original dataset.

7. CONCLUSIONS

This paper has presented a technique of Naïve-Bayes and K-nearest neighbors (KNN) to predict chronic kidney disease (CKD) using machine learning. By taking into consideration of 1500 instances of dataset with all 24 attributes, the experimental results are obtained. The results conclude that Naïve-Bayes gives the highest accuracy with 91.549295774647% as that of with K-nearest neighbors which has given an accuracy of 87.323943661971%. Through this we can conclude that Naïve Bayes has highest sensitivity after training and testing by the proposed method. Therefore, it can be concluded that classifier is appropriated for predicting the chronic kidney disease. The main scope here is we have used 1500 instance of dataset with 24 different parameters which is more helpful to predict the disease earlier with more accuracy and

is more effective. Hence Naïve Bayes classifier is more suitable for earlier prediction of CKD than K-nearest neighbors in our experiment. And the GFR technique for stage prediction.

8. REFERENCES

[1] Anusorn Charleonnan, Thipwan Fufaung, Tippawan Niyomwong, Wandee Chokchueypattanakit, Sathit Suwannawach, Nitat Ninchawee, "Predictive Analytics for Chronic Kidney Disease Using Machine Learning Techniques" The 2016 Management and Innovation Technology International Conference (MITiCON-2016)

[2] Helmie Arif ,Indra Malik, Nurdin 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)

[3] Devika R, Sai Vaishnavi Avilala, V. Subramaniyaswamy "Comparative Study of Classifier for Chronic Kidney Disease prediction using Naive Bayes, KNN and Random Forest" Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019) IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4

[4] Lambodar Jena, Ramakrushna Swain "Work-in-Progress: Chronic Disease Risk Prediction using Distributed Machine Learning Classifiers" 2017 International Conference on Information Technology

[5] Ahmed J. Aljaaf1, 2, Dhiya Al-Jumeily2, Hussein M. Haglan1, Mohamed Alloghani3, Thar Baker2, Abir J. Hussain2, andJamila Mustafina4 "Early Prediction of Chronic Kidney Disease Using Machine Learning Supported by Predictive Analytics" @ljmu.ac.uk;M.AlLawghani@2014.ljmu.ac.uk;

[6] Gunarathne W.H.S.D, Perera K.D.M, Kahandawaarachchi K.A.D.C.P "Performance Evaluation on Machine Learning Classification Techniques for Disease Classification and Forecasting through Data Analytics for Chronic Kidney Disease (CKD)" 2017 IEEE 17th International Conference on Bioinformatics and Bioengineering

[7] AKM Shahariar Azad Rabby, Rezwana Mamata, Monira Akter Laboni, Ohidujjaman, Sheikh Abujar. "Machine Learning Applied to Kidney Disease Prediction: Comparison Study" 10th ICCCNT 2019 July 6-8, 2019, IIT - Kanpur Kanpur, India

[8] Nagaraju C and Parthasarathy S S "Quality Measure for Information Hiding in Medical Images" International Journal of Computer Applications, IJCA-Volume 65– No.11, March 2013

[9] Dr. Nagaraju C, Sangana Manasa Durga and Rachana N (2017) "Ascertainment of Lung Cancer at An Early Stage", Aug 2017 IJSRCSEIT, Volume 2, Issue 4, ISSN : 2456-3307.

[10] Nagaraju C and Parthasarathy S S (2013) "Quality Measure for Information Hiding in Medical Images" International Journal of Computer Applications, IJCA-Volume 65– No.11, ISSN- 0975 – 8887, March 2013.

BIOGRAPHIES

Dr. C Nagaraju completed his B.E and M-Tech from SJCE, Mysore, and Karnataka, India. He has completed his Ph.D. at PET Research Centre, Mandya under the guidance of **Dr. S S Parthasarathy**. He has 15 years of teaching experience and currently working as Assistant Professor, in department of ECE at The National Institute of Engineering, an autonomous institution, Mysore, Karnataka, India. He has published more than 30 research publications in various National, International journals and Conference proceedings. His research interest includes Image Processing and Medical Image analysis.

Varun B completed his B.E from Vidya vardhaka college of engineering (VVCE), Mysore, Karnataka, India. He is currently pursuing his Masters (M. Tech) in Network and internet engineering, dept of ECE, National Institute of Engineering (NIE), an autonomous institution Mysore, Karnataka, India.