EARLY DETECTION OF CHRONIC KIDNEY DISEASE USING MACHINE LEARNING

Abstract:

Chronic Kidney Disease is a disorder that disables normal kidney function. The WHO has shown that CKD is a serious disease, ranked as one of the top twenty causes of death. It is recognized that 2 million people worldwide suffer from kidney failure and the number of patients diagnosed with CDK continues to expand at a rate of 5-7% annually. Late diagnosis of this disease is a life-threatening problem, which often occurs in remote areas due to the lack of specialized medical personnel, in addition to the high cost of diagnosis. This paper aims at early detection of CDK using Machine Learning Algorithms Artificial Neural Network, Support Vector Machine, and k-Nearest Neighbor. The importance of Machine Learning is reflected in the importance of identifying these typically fatal ailments. This study looks at a data set consisting of 400 samples and 13 features. The three classification techniques were evaluated by applying them to the data. The results show that the ANN classifier achieved the best accuracy at 99.2%.

PAPER 1: A Deep Neural Network for Early Detection and Prediction of Chronic Kidney Disease [Vijendra Singh, Vijayan K. Asari and Rajkumar Rajasekaran].

Diabetes and high blood pressure are the primary causes of Chronic Kidney Disease (CKD). Glomerular Filtration Rate (GFR) and kidney damage markers are used by researchers around the world to identify CKD as a condition that leads to reduced renal function over time. A person with CKD has a higher chance of dying young. Doctors face a difficult task in diagnosing the different diseases linked to CKD at an early stage in order to prevent the disease. This research presents a novel deep learning model for the early detection and prediction of CKD. This research objectives to create a deep neural network and compare its performance to that of other contemporary machine learning techniques. In tests, the average of the associated features was used to replace all missing values in the database. After that, the neural network's optimum parameters were fixed by establishing the parameters and running multiple trials. The foremost important features were selected by Recursive Feature Elimination (RFE). Haemoglobin, Specific Gravity, Serum Creatinine, Red Blood Cell Count, Albumin, Packed Cell Volume, and Hypertension were found as key features in the RFE. Selected features were passed to machine learning models for classification purposes. The proposed Deep neural model outperformed the other four classifiers (Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Logistic regression, Random Forest, and Naive Bayes classifier) by achieving 100% accuracy. The proposed approach could be a useful tool for nephrologists in detecting CKD.

PAPER 2: Detection of Chronic Kidney Disease using Machine Learning Algorithms with Least Number of Predictors[Marwa Almasound,Tomas E Ward],Information System Department | College of Computer and Information Science |(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 8, 2019.

Chronic kidney disease (CKD) is one of the most critical health problems due to its increasing prevalence. In this paper, we aim to test the ability of machine learning algorithms for the prediction of chronic kidney disease using the smallest subset of features. Several statistical tests have been done to remove redundant features such as the ANOVA test, the Pearson's correlation, and the Cramer's V test. Logistic regression, support vector machines, random forest, and gradient boosting algorithms have been trained and tested using 10-fold cross-validation. We achieve an accuracy of 99.1 according to F1- measure from Gradient Boosting classifier. Also, we found that haemoglobin has higher importance for both random forest and Gradient boosting in detecting CKD. Finally, our results are among the highest compared to previous studies but with less number of features reached so far. Hence, we can detect CKD at only $26.65 by performing three simple tests.

PAPER 3: A Literature Review on Kidney Disease Prediction using Data Mining Classification Technique[Suman Bala, Krishan Kumar], Department of Computer Science& Engineering, JNTU Hyderabad, India | IJCSMC, Vol. 3, Issue. 7, July 2014, pg.960 – 967.

Chronic Kidney Disease (CKD) is a type of chronic disease which means it happens slowly over a period of time and persists for a long time thereafter. It is deadly at its end stage and will only be cured by kidney replacement or regular dialysis which is an artificial filtering mechanism. It is important to identify CKD at the early stage so that necessary treatments can be provided to prevent or cure the disease. The main focus in this paper is on the classification techniques, that is, tree-based decision tree, random forest, and logistic regression has been analysed. Different measure has been used for comparison between algorithms for the dataset collected from standard UCI repository. That over a period of time and persists for a long time thereafter. It is deadly at its end stage and will only be cured by kidney replacement or regular dialysis which is an artificial filtering mechanism. It is important to identify CKD at the early stage so that necessary treatments can be provided to prevent or cure the disease.

PAPER 4: A Hybrid Parallel Classification Model for the Diagnosis of Chronic Kidney Disease [Vijendra Singh, Divya Jain],School of Computer Science, University of Petroleum and Energy Studies, Dehradun|Computer Science and Engineering, The NorthCap University, Gurugram.

Chronic Kidney Disease (CKD) has become a prevalent disease nowadays, affecting people globally around the world. Accurate prediction of CKD progression over time is essential for reducing its associated mortality and morbidity rates. This paper proposes a fast, novel hybrid approach to diagnose Chronic Renal Disease. The proposed approach is based on the optimization of SVM classifier with

the hybridized dimensionality reduction approach to identify the most informative parameters for CKD diagnosis. It handles the selection of features through two steps. The first one is a filter-based approach using ReliefF method to assign weights and ranks to each feature of the dataset. The second step is the dimensionality reduction of the best-selected subset by means of PCA, a feature extraction technique. For faster execution of datasets, simultaneous execution on multiple processors is employed. The proposed model achieved the highest prediction accuracy of 92.5% on the clinical CKD dataset compared to existing methods - 'CFS+SVM' (60.45%), 'ReliefF + SVM' (86%), 'MIFS + SVM' (56.72%), 'ReliefF + CFS + SVM' (54.37). The proposed work is also examined on the benchmarked Chronic Kidney Disease Dataset and achieved classification accuracy of 98.5% compared to the accuracy with other methods -'CFS+SVM' (92.7%), 'ReliefF + SVM' (89.6%), 'MIFS + SVM' (94.7%). The experimental outcomes positively demonstrate that the proposed hybridized model is effective in undertaking medical data classification tasks and is, therefore, a promising tool for the diagnosis of CKD patients. The proposed approach is statistically validated with the Friedman test with significant results compared to other techniques. The proposed approach also executes in the least time with improved prediction accuracy and competes with and even outperforms other methods in the literature.

PAPER 5: Development of a Graphical User Interface Software for The Prediction of Chronic Kidney Disease[S.C. Nwaneri1, H.C. Ugo],Nigerian Journal of Technology (NIJOTECH) Vol. 41, No. 1, January, 2022, pp.175 –183.

Chronic Kidney Disease (CKD) is a severe kidney damage that is difficult to diagnose at the early stages due to the absence of clear symptoms.. This

study is designed to develop a user-friendly web-based graphical user interface (GUI) software for the prediction of CKD using artificial neural networks (ANNs). The model was developed using Python programming language and trained with 1200 instances of CKD datasets obtained from the University of California Irvine (UCI) machine learning repository. This dataset was split into 80% for training and 20% for testing achieved through an iterative process. A GUI software was developed based on the model using Django, an open-source python web development framework. The model achieved an accuracy of 95.83%, a precision of 100%, a specificity of 100%, and a sensitivity of 89.80%. The GUI software was effectively used to predict CKD and could be of immense benefit as a point of care application for early CKD prediction.

PAPER 6: Analysis and Prediction of Chronic Kidney Disease using Machine Learning Classification approaches [Abhimanyu Agarwala , Asfar Sharief,Faaiz Ahmed] ,Department of Computer science and Engineering CMR Institute of Technology,Bangalore-37.

Chronic Kidney Disease (CKD) is one of the deadliest diseases that slowly damages human kidney. The disease remains undetected in its early stage and the patients can only realize the severity of the disease when it gets advanced. Hence, detecting such disease at earlier stage is a key challenge now. Machine Learning is one of the emerging field used in the health sectors for the diagnosis of different diseases. In this paper, we compute, analyse and compare between Machine Learning classification approaches to determine which classification approach is the optimal for the prediction of CKD. K-Nearest Neighbor Classifier, Decision Tree Classifier, GaussianNB, Logical Regression and Artificial Neural Network (ANN)

are some renowned machine learning methods which were selected to train the model and based on these results, we can compare and determine which among the following Machine Learning Methods can predict the possibility of CKD at the most accurate level.From this comparative analysis, Logistic Regression is found to be the best approach to predict CKD.

PAPER 7: Data Mining Performance in Identifying the Risk Factors of Early Arteriovenous Fistula Failure in Haemodialysis Patients[Morteza Khavanin Zadeh; Mohammad Rezapour; Mohammad Mehdi Sepehri], Volume 2, Issue 1, March 2013, Pages 49-54.

Arteriovenous fistula is a popular vascular access method for surgical treatment of haemodialysis patients. The method, however, is associated with a high rate of early failure varying in the range of 20-60%. Predicting early Arteriovenous fistula failure and its risk factors can help reduce its incidence, its hospitalization rate, and associated costs. In this study, we examined performance of data mining in the prediction of early AVF failure and identification of its risk factors. The data of 193 patients who underwent haemodialysis in Hasheminejad Kidney Centre were explored. Eight common attributes of the patients including age, sex, hypertension level, Diabetes Mellitus state, haemoglobin level, smoking behaviour, location of Arteriovenous fistula, and thrombosis state were used in the machine learning process. Two learning operators including W-Simple Cart and WJ48 tree were used in data mining process. Smoking was identified as a factor influencing the relationship between the outcome of vascular access surgery and haemoglobin level. Prediction accuracy varied within the range of 69.15-85.11%. According to our results smoking is a crucial risk factor for early Arteriovenous fistula failure, even

at normal levels of haemoglobin. Our results provide further supports for the notion that data mining can help medical decision-making process by deciphering the complex interactions between various biological variables and translating the hidden patterns in data into detailed decision-making criteria.