



AN IMPROVED DEEP NEURAL NETWORK FOR CHRONIC KIDNEY DISEASE PREDICTION

Sankara Krishnan G¹, Vignesh M², Arul Pandian M³, Mrs.D.Thamarai Selvi⁴

Author Affiliations

¹4th Year Student, Department of
Computer Science Engineering, National
Engineering College, Kovilpatti.

²4th Year Student, Department of
Computer Science Engineering, National
Engineering College, Kovilpatti.

³4th Year Student, Department of
Computer Science Engineering, National
Engineering College, Kovilpatti.

⁴Assistant Professor, Department of
Computer Science Engineering, National
Engineering College, Kovilpatti.

ABSTRACT: Chronic kidney Disease (CKD) means your kidneys are damaged and not filtering your blood the way it should. The primary role of kidneys is to filter extra water and waste from The associate editor coordinating the review of this manuscript and approving it for publication was Haruna Chiroma. your blood to produce urine and if the person has suffered from CKD, it means that wastes are collected in the body. This disease is chronic because of the damage gradually over a long period. It is flatterring a common disease worldwide. Due to CKD may have some health troubles. There are many causes of CKD like diabetes, high blood pressure, and heart disease. Prediction of Chronic Kidney Disease depends on age and gender. If your kidney is not working, then you may notice one or more symptoms like abdominal pain, back pain, diarrhea, fever, nosebleeds, rash, vomiting In our project, we are going to improve the Deep neural network for detecting chronic kidney disease by increasing the number of hidden layers in it. So that we can able to get high accuracy during classification.

KEYWORDS: Deep neural network, hidden layers, CKD.

I. INTRODUCTION

Prediction

“Prediction” refers to the output of an algorithm after it has been trained on a historical data set and applied to new data when forecasting the likelihood of a particular outcome, such as whether or not a customer will churn in 30 days. The algorithm will generate probable values for an unknown variable for each record in the new data, allowing the model builder to identify what that value will most likely be. The word “prediction” can be misleading. In some cases, it really does mean that you are predicting a future outcome, such as when you’re using machine learning to determine the next best action in a marketing campaign. Other times, though, the “prediction” has to do with, for example, whether or not a transaction that already occurred was fraudulent. In that case, the transaction already happened, but you’re making an educated guess about whether or not it was legitimate, allowing you to take the appropriate action.

Chronic Kidney Disease (CKD)

Chronic Kidney Disease (CKD) means your kidneys are damaged and not filtering your blood the way it should. The primary role of kidneys is to filter extra water and waste from your blood to produce urine and if the person has suffered from CKD, it means that wastes are collected in the body. This disease is chronic because of the damage gradually over a long period. It is becoming a common disease worldwide. Due to CKD may have some health troubles. There are many causes of CKD like diabetes, high blood pressure, and heart disease. Along with these critical diseases, CKD also depends on age and gender. If your kidney is not working, then you may notice one or more symptoms like abdominal pain, back pain, diarrhea, fever, nosebleeds, rash, and vomiting. There are two main diseases of CKD: (i) diabetes and (ii) high blood pressure. So that controlling of these two diseases is the prevention of CKD. Usually, CKD does not give any sign till the kidney is damaged badly. CKD is being increased rapidly as per the studies hospitalization cases increase by 6.23 percent per year but the global mortality rate remains high [4]. There are few diagnostic tests to check the condition of CKD: (i) estimated glomerular filtration rate (eGFR) (ii) urine test (iii) blood pressure.

EGFR

EGFR value shows how your kidney cleanses the blood. If your eGFR value is greater than 90, that means the kidney is normal. If the eGFR value is less than 60, that means you have CKD.

URINE TEST

The doctor also asks for a urine test for kidney functionality because kidneys make urine. If the urine contains blood and protein [6], that means your kidney is not working properly.

BLOOD PRESSURE

The doctor measures blood pressure as Blood pressure range shows how your heart is pumping blood. If the eGFR value reaches less than 15, that means the patient has end-stage kidney disease. At this point, there are only available treatments: (i) dialysis and (ii) kidney transplant. A patient's life after dialysis depends on such factors as age, gender, frequency and duration of dialysis, physical movement of the body, and mental health. If dialysis is not possible, the doctor has only one solution, i.e., kidney transplantation. However, it is extremely expensive. Therefore, it is critical noteworthy in early recognition, monitoring, and handling of the disease. It is essential to predict the striding of CKD with appropriate accuracy due to its dynamic and secretive nature in the early stages and patient abnormality.

II. LITERATURE SURVEY

1. “Prediction of Chronic Kidney Disease - A Machine Learning Perspective”, Volume 10, 2021

Authors: PANKAJ CHITTORA 1, SANDEEP CHAURASIA1, (Senior Member, IEEE), PRASUN CHAKRABARTI2,3

Chronic Kidney Disease is one of the most critical illnesses nowadays and proper diagnosis is required as soon as possible. Machine learning technique has become reliable for medical treatment. With the help of machine learning classifier algorithms, the doctor can detect the disease on time. For this perspective, Chronic Kidney Disease prediction has been discussed in this article. Chronic Kidney Disease data set has been taken from the UCI repository. Seven classifier algorithms have been applied in this research

such as artificial neural network, C5.0, Chi-square Automatic interaction detector, logistic regression, linear support vector machine with penalty L1 & with penalty L2, and random tree. The important feature selection technique was also applied to the data set. Their future work is to use Hybrid Machine learning techniques to improve prediction accuracy.

2. “An Empirical Evaluation of Machine Learning Techniques for Chronic Kidney Disease Prophecy”, Volume8, 2020

Authors: BILAL KHAN 1, RASHID NASEEM 2, FAZAL MUHAMMAD 3 , GHULAM ABBAS 4, (Senior Member, IEEE), AND SUNGHWAN KIM

In this paper, we employ experiential analysis of ML techniques for classifying the kidney patient data set as CKD or NOTCKD. Seven ML techniques together with NBTree, J48, Support Vector Machine, Logistic Regression, Multilayer Perceptron, Naïve Bayes, and Composite Hypercube on Iterated Random Projection (CHIRP) are utilized and assessed using distinctive evaluation measures such as mean absolute error (MAE), root means squared error (RMSE), relative absolute error (RAE), root relative squared error (RRSE), recall, precision, F-measure and accuracy. Their Future work is to implement this project using a Deep Neural network.

3. “A Machine Learning Methodology for Diagnosing Chronic Kidney Disease”, Volume 57, 2019

Authors: JIONGMING QIN 1, LIN CHEN 2, YUHUA LIU 1, CHUANJUN LIU 2, CHANGHAO FENG 1, AND BIN CHEN 1

In this study, we propose a machine learning methodology for diagnosing CKD. The CKD data set was obtained from the University of California Irvine (UCI) machine learning repository, which has a large number of missing values. KNN imputation was used to fill in the missing values, which selects several complete samples with the most similar measurements to process the missing data for each incomplete sample. Missing values are usually seen in real-life medical situations because patients may miss some measurements for various reasons. After effectively filling out the incomplete data set, six machine learning algorithms (logistic regression, random forest, support vector machine, k-nearest neighbor, naive Bayes classifier, and feed-forward neural network) were used to establish models. Among these machine learning models, random forest

achieved the best performance with 99.75% diagnosis accuracy. Their future work is to use hybrid techniques.

4. “Clinically Applicable Machine Learning Approaches to Identify Attributes of Chronic Kidney Disease (CKD) for Use in Low-Cost Diagnostic Screening”, Volume 9, 2021

Authors: MD. RASHED-AL-MAHFUZ 1, ABEDUL HAQUE 2, AKM AZAD3

In this study, we developed machine learning models using selective key pathological categories to identify clinical test attributes that will aid in accurate early diagnosis of CKD. Such an approach will save time and costs for diagnostic screening. We have also evaluated the performance of several classifiers with k-fold cross-validation on optimized datasets derived using these selected clinical test attributes. Results: Our results suggest that the optimized datasets with important attributes perform well in diagnosis of CKD using our proposed machine learning models. Their future work is increase the input attributes and then test it.

5. “Cluster Analysis of Mixed and Missing Chronic Kidney Disease Data in KwaZulu-Natal Province, South Africa”, Volume 9, 2021

Authors: HUI ZHANG 1, YURONG CHEN1, YANAN SONG1, ZHENLIN XIONG1, YIMIN YANG 2, (Member, IEEE), AND Q. M. JONATHAN WU 3

The CT scan image is one of the most useful tools for diagnosing and locating lesions in the kidney. It can provide precise information about the location and size of lesions in many medical applications. Manual and traditional medical testings are labor-consuming and time-costing. Nowadays, detecting lesions in CT automatically is an integral assignment to the paramount importance of clinical diagnosis. Computer-aided diagnosis (CAD) is needed to develop and improve medical testing efficiency. However, it is still a tremendous challenge to the extant low precision and incomplete detection algorithm. In this paper, we proposed a lesion detection tool using multi intersection over union (IOU) threshold based on morphological cascade conventional neural networks (CNNs). Their future work is to use Resnet and Unet architecture.

6. “Kidney Segmentation in Renal Magnetic Resonance Imaging - Current Status and Prospects”, Volume 16 2021

Authors:FRANK G. ZÖLLNER 1 , (Senior Member, IEEE), MAREK KOCIŃSKI 2,4 , LAURA HANSEN1 , ALENA-KATHRIN GOLLA 1 , AMIRA ŠERIFOVIĆ TRBALIĆ 3 , ARVID LUNDERVOLD2

Magnetic resonance imaging has achieved an increasingly important role in the clinical work-up of renal diseases such chronic kidney disease (CKD). A large panel of parameters have been proposed to diagnose CKD among them total kidney volume (TKV) which recently qualified as biomarker. Volume estimation in renal MRI is based on image segmentation of the kidney and/or its compartments. Beyond volume estimation renal segmentation supports also the quantification of other MR-based parameters such as perfusion or filtration. The aim of the present article is to discuss the recent existing literature on renal image segmentation techniques and show today's limitations of the proposed techniques that might hinder clinical translation. We also provide pointers to open source software related to renal image segmentation. Their future work is to use RCNN and UNET for segmentation.

III. PROPOSED SYSTEM

In the proposed system we use a Deep learning model which is called as Deep Neural Network which is suitable for Accurate Prediction. By using DNN, we can predict Chronic Kidney Disease with more than 95% of accuracy. In the DNN we have more hidden layers and hence its accuracy also high.

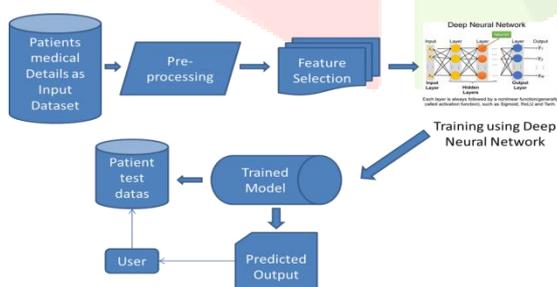


Fig 1. Proposed Architecture

DATA COLLECTION

We collected the Kidney disease data set from the Kaggle website. The data was taken over 2 months in India with 25 features (eg, red blood cell count, white blood cell count, etc). The target is the 'classification', which is either 'CKD' or 'notckd' - **ckd=chronic kidney disease**. There are 400 rows The data needs cleaning: in that it has NaNs and the numeric features need to be forced to floats. Basically, we were instructed to get rid of

ALL ROWS with Nans, with no threshold - meaning, any row that has even one NaN, gets deleted. Part 1: We are asked to choose 3 features (*bgr, rc, wc*), visualize them, then run the PCA with *n_components=2*. the PCA is to be **run twice**: one with **no scaling** and the second run **WITH** scaling. And this is where my issue starts ... **in that after scaling I can hardly see any difference!** I will stop here for now till I get feedback and then move to Part 2.



Fig 2. Data Collection

DATA PRE-PROCESSING

At first, the data set is fetched by using the pandas library and then we save the data s inside a pandas data frame, At first, this data set consists of lots of null values, then we drop all the null values into 0 because our Deep learning model cannot able to process null values.

DEEP NEURAL NETWORK

Nodes are little parts of the system, and they are like neurons of the human brain. When a stimulus hits them, a process takes place in these nodes. Some of them are connected and marked, and some are not, but in general, nodes are grouped into layers. The system must process layers of data between the input and output to solve a task. The more layers it has to process to get the result, the deeper the network is considered. There is a concept of Credit Assignment Path (CAP) which means the number of such layers needed for the system to complete the task. The neural network is deep if the CAP index is more than two. A deep neural network is beneficial when you need to replace human labor with autonomous work without compromising its efficiency. Deep neural network usage can find various applications in real life. For example,

A Chinese company Sense-time created a system of automatic face recognition system to identify criminals, which uses real-time cameras to find an offender in the crowd. Nowadays, it has become a popular practice in police and other governmental entities.

outputs accomplish the task, such as recognizing an object in an image.

Perceptron Mathematical Model

In machine learning, the perceptron is an algorithm for supervised learning of binary classifiers. A binary classifier is a function which can decide whether or not an input, represented by a vector of numbers, belongs to some specific class. It is a type of linear classifier, i.e. a classification algorithm that makes its predictions based on a linear predictor function combining a set of weights with the feature vector.

$$f(x) = b_0 + \sum_{i=0}^n (b_i * W_i)$$

$b_i =$
inputs

$W_i =$
Weights

$b_0 =$
bias

A multi-layer perceptron (MLP) is a class of feed-forward artificial neural network (ANN). The term MLP is used ambiguously, sometimes loosely to *any* feed-forward ANN, sometimes strictly to refer to networks composed of multiple layers of perceptrons (with threshold activation). Multi-layer perceptrons are sometimes colloquially referred to as "vanilla" neural networks, especially when they have a single hidden layer. An MLP consists of at least three layers of nodes: an input layer, a hidden layer, and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called back propagation for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable.

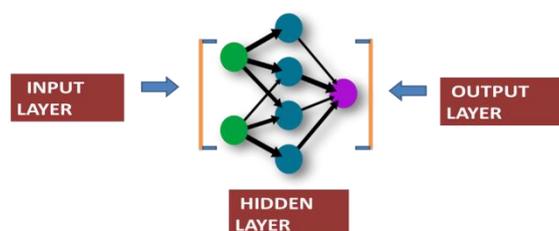


Fig 6. Multiple Perceptron Networks

Back-Propagation

In machine learning, back propagation (backprop, BP) is a widely used algorithm for training feed-forward neural networks. Generalizations of back propagation exist for other artificial neural

networks (ANNs), and functions generally. These classes of algorithms are all referred to generically as "back propagation". In fitting a neural network, back propagation computes the gradient of the loss function concerning the weights of the network for a single input-output example, and does so efficiently, unlike a naive direct computation of the gradient concerning each weight individually. This efficiency makes it feasible to use gradient methods for training multi-layer networks, updating weights to minimize loss; gradient descent, or variants such as stochastic gradient descent, are commonly used. The back propagation algorithm works by computing the gradient of the loss function concerning each weight by the chain rule, computing the gradient one layer at a time, iterating backward from the last layer to avoid redundant calculations of intermediate terms in the chain rule; this is an example of dynamic programming.

Now you might be suddenly bogged with the question, why Python? According to the Institute of Electrical and Electronics Engineers (IEEE), 2016 ranking Python ranked third after C and Java. As per Indeed.com's data for 2016, the Python job market search ranked fifth. All the data points to the ever-rising demand in the job market for Python. It's a cool language if you want to learn just for fun or if you want to build your career around Python, you will adore the language. At the school level, many schools have started including Python programming for kids. With new technologies taking the market by surprise Python has been playing a dominant role. Whether it is a cloud platform, mobile app development, Big Data, IoT with Raspberry Pi, or the new Blockchain technology, Python is being seen as a niche language platform to develop and deliver a scalable and robust applications.

Python programs can run on any platform, you can carry code created on a Windows machine and run it on Mac or Linux. Python has an inbuilt large library with prebuilt and portable functionality, also known as the standard library. Python is an expressive language. Python is free and open-source. Python code is about one-third of the size of equivalent C++ and Java code. Python can be both dynamically and strongly typed--dynamically typed means it is a type of variable that is interpreted at runtime, which means, in Python, there is no need to define the type (int or float) of the variable.

IV. Chronic Kidney Disease Prediction

In the proposed system we use a Deep learning model which is called as Deep Neural Network which is suitable for Accurate Prediction. By using DNN, we can predict Chronic Kidney Disease with more than 95% of accuracy. In the DNN we have more hidden layers and hence its accuracy also high.

DEEP LEARNING

Deep learning (also known as deep structured learning) is part of a broader family of machine learning methods based on artificial neural networks with representation learning. Learning can be supervised, semi-supervised or unsupervised. Deep learning architectures such as deep neural networks, deep belief networks, recurrent neural networks, and convolution neural networks have been applied to fields including computer vision, machine vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, bioinformatics, drug design, medical image analysis, material inspection, and board game programs, where they have produced results comparable to and in some cases surpassing human expert performance. The adjective "deep" in deep learning comes from the use of multiple layers in the network. Early work showed that a linear perceptron cannot be a universal classifier, and then that a network with a non-polynomial activation function with one hidden layer of unbounded width can on the other hand so be. Deep learning is a modern variation that is concerned with an unbounded number of layers of bounded size, which permits practical application and optimized implementation while retaining theoretical universality under mild conditions. In deep learning the layers are also permitted to be heterogeneous and to deviate widely from biologically informed connectionist models, for the sake of efficiency, trainability, and understandability, whence the "structured" part.

DEEP NEURAL NETWORK

Nodes are little parts of the system, and they are like neurons of the human brain. When a stimulus hits them, a process takes place in these nodes. Some of them are connected and marked, and some are not, but in general, nodes are grouped into layers.

The system must process layers of data between the input and output to solve a task. The more layers it has to process to get the result, the deeper the network is considered. There is a

concept of Credit Assignment Path (CAP) which means the number of such layers needed for the system to complete the task. The neural network is deep if the CAP index is more than two.

A deep neural network is beneficial when you need to replace human labor with autonomous work without compromising its efficiency. The deep neural network usage can find various applications in real life. For example, a Chinese company SenseTime created a system of automatic face recognition system to identify criminals, which uses real-time cameras to find an offender in the crowd. Nowadays, it has become a popular practice in police and other governmental entities.



Fig 7. Chronic Kidney Disease Prediction

V. CONCLUSION

In our project, we used Deep Neural Network for predicting Chronic kidney disease. It is one of the deep learning techniques which can be able to train and predict Kidney disease based upon input data. We used the Chronic kidney disease data set which holds over 24 features for training purposes. After training, we predict chronic kidney disease by using test data's using DNN. Our Model Archives have more than 96% accuracy during testing and training. The predicted results by DNN are accurate and stable, its patterns are also matched with the existing data set patterns. And hence our model is perfectly trained and it can able to predict kidney disease with high stability. Our future work is to improve accuracy by using hybrid machine learning techniques.

VI. REFERENCES

1. Z. Chen, Z. Zhang, R. Zhu, Y. Xiang, and P. B. Harrington, "Diagnosis of patients with chronic kidney disease by using two fuzzy classifiers," *Chemometrics Intell. Lab. Syst.*, vol. 153, pp. 140145, Apr. 2016.
2. A. Subasi, E. Alickovic, and J. Kevric, "Diagnosis of chronic kidney disease by using random forest," in *Proc. Int. Conf. Med. Biol. Eng.*, Mar. 2017, pp. 589594.
3. L. Zhang, "Prevalence of chronic kidney disease in China: A crosssectional survey," *Lancet*, vol. 379, pp. 815822, Mar. 2012.
4. A. Singh, G. Nadkarni, O. Gottesman, S. B. Ellis, E. P. Bottinger, and J. V. Guttag, "Incorporating temporal EHR data in predictive models for risk stratification of renal function deterioration," *J. Biomed. Informat.*, vol. 53, pp. 220228, Feb. 2015.
5. A. M. Cueto-Manzano, L. Cortés-Sanabria, H. R. Martínez-Ramírez, E. Rojas-Campos, B. Gómez-Navarro, and M. Castellero-Manzano, "Prevalence of chronic kidney disease in an adult population," *Arch. Med. Res.*, vol. 45, no. 6, pp. 507513, Aug. 2014.
6. H. Polat, H. D. Mehr, and A. Cetin, "Diagnosis of chronic kidney disease based on support vector machine by feature selection methods," *J. Med. Syst.*, vol. 41, no. 4, p. 55, Apr. 2017.
7. C. Barbieri, F. Mari, A. Stopper, E. Gatti, P. Escandell-Montero, J. M. Martínez-Martínez, and J. D. Martín-Guerrero, "A new machine learning approach for predicting the response to anemia treatment in a large cohort of end stage renal disease patients undergoing dialysis," *Comput. Biol. Med.*, vol. 61, pp. 5661, Jun. 2015.

