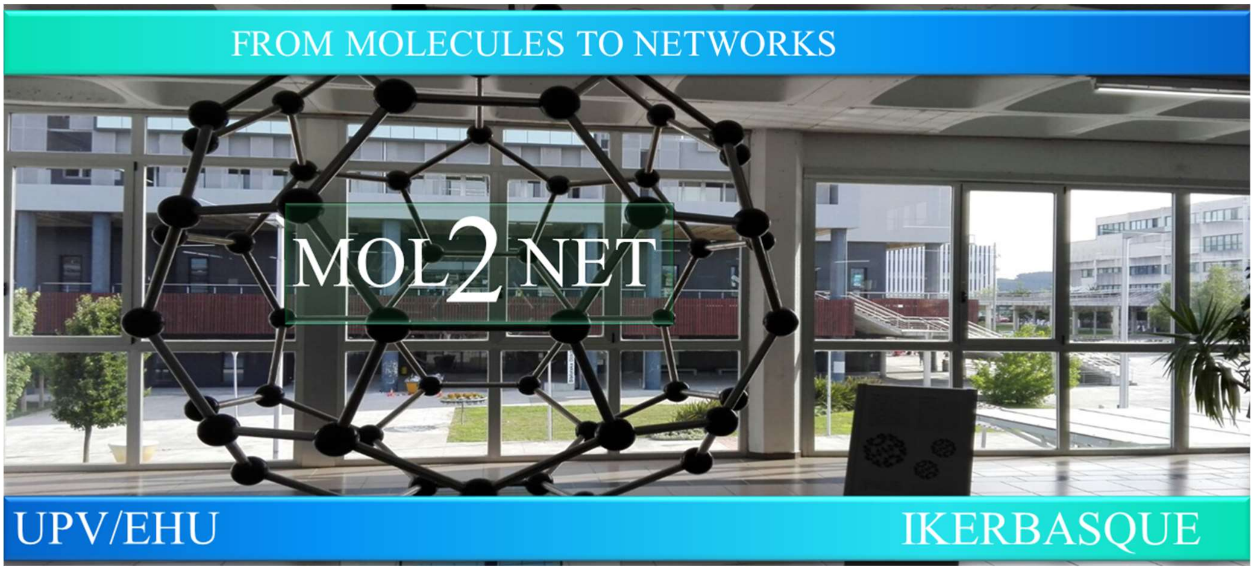
MOL2NET'22, Conference on Molecular, Biomedical & Computational Sciences and Engineering, 8th ed.



Machine Learning Based Classification of Chronic Kidney Disease Using CT Scan Images

Aqib Ali [*, a], *Samreen Naeem* [a], *and Muhammad Zubair* [a],

[a] *College of Automation, Southeast University Nanjing China.*

.

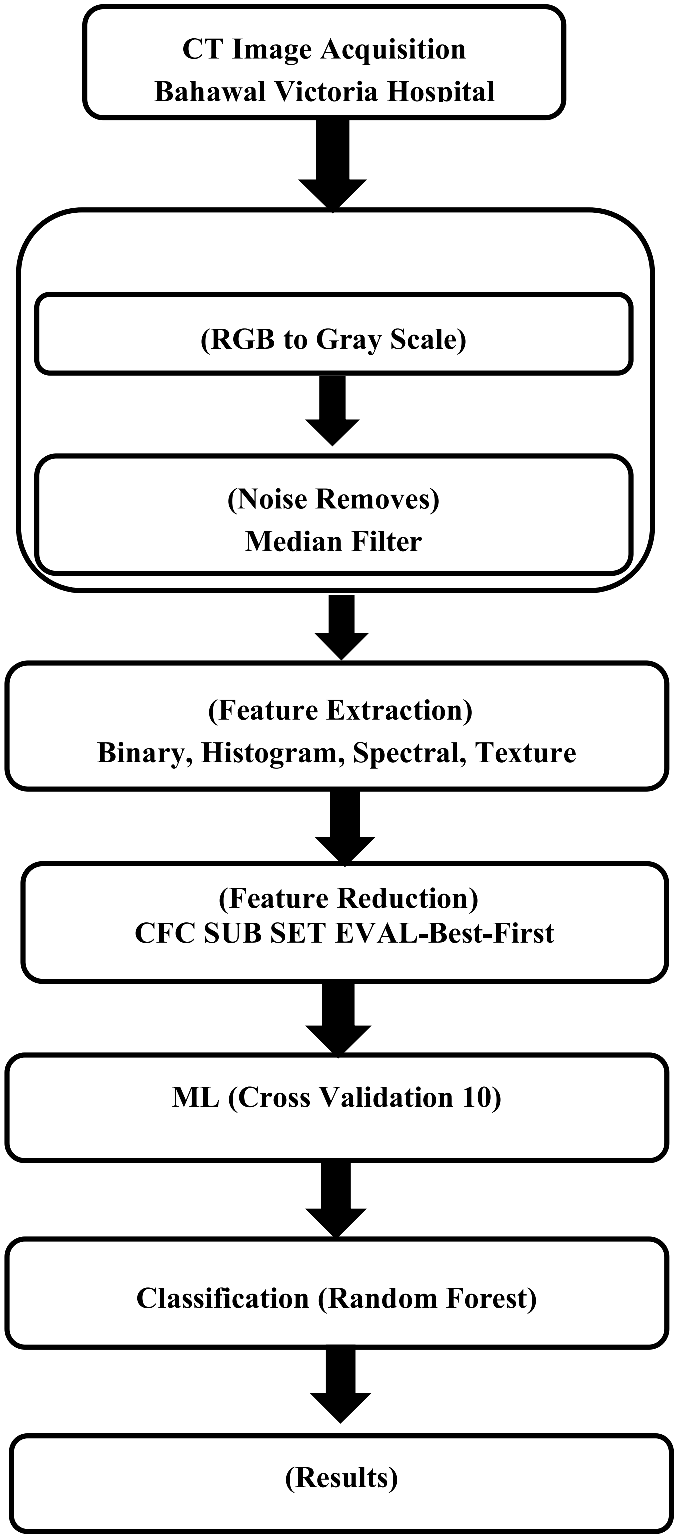
* *Corresponding author:* aqibcsit@gmail.com

.

Abstract

Imaging techniques such as CT scans have found widespread application in kidney diagnosis. These imaging techniques can estimate kidney size, shape, and position; provide information about kidney function; and assist in diagnosing structural abnormalities such as cysts, stones, and infections. However, different operators have different levels of success when it comes to using CT scans to diagnose renal conditions. The images can be interpreted in various ways due to factors such as the abilities and experiences of the operators, variances in how individuals see the images, and changes in the characteristics utilized for diagnosis. The detection of chronic renal disease might be improved with automated approaches and computer-aided diagnosis systems; however, research into these methods has been limited. According to the findings of this research, the Random Forest classifier has the highest level of accuracy (96.33%) among the various Machine Learning classifiers. As a result, the researchers concluded that chronic renal disease might have been caused by its acquisition. The outcomes of this study indicate that further research should be conducted. Suppose these suggested algorithms were integrated into the existing methods for diagnosing chronic renal disease. In that case, it might be possible to increase the accuracy of present diagnostic procedures while simultaneously lowering the need for human involvement and operator dependence.

Graphical Abstract



Introduction

Kidney chronic disease is a fatal and life-threatening disease for men and women. The main reason is smoking and drinking alcohol [1]. The survival rate from this disease is meager, but one thing can save the patient if the detection of chronic disease is done at early stages. Usually, the human body grows slowly, but abnormal cells grow fast without any ratio. This abnormal division process of cells is called chronic disease. Abnormal cells are divided abnormally or can damage other related cells, so early detection and control are challenging for doctors [2]. National Cancer Institute (NCI) 2019 reported that 224210 kidney chronic disease cases were found, in which 159260 people died due to chronic kidney disease [3]. It is challenging to detect chronic kidney disease early on with the naked eye, so a particular technique or method must be needed to diagnose it accurately with early symptoms. So, Computer Tomography (CT) is one of the most popular ways to detect chronic kidney disease by using CT images of the patient. Computer Tomography (CT) images are low in cost and give precise results [4].

According to survey, images are used as input, and this input is used for enhancement. After enhancing the image, it is segmented and further used for feature extraction, so the result extracted is helpful for chronic disease detection. So, like other systems, it is also beneficial for the early detection of chronic kidney disease [5]**.** Kidney chronic disease is a dangerous and death-causing disease worldwide. According to the World Health Organization (WHO), 64 million people worldwide have been infected, and the death rate can lead to one-third in 2030. According to another report by WHO, 1.5 million people died in 2012 due to chronic kidney disease, which is becoming life-threatening to human beings [6]. This disease can only be engulfed by detecting accurate results, or proper medical treatment can lead the patient to a healthy, safe life. So, we can detect it early by using CT scan images and improve the survival rate [5]. We are trying to detect chronic kidney disease by collecting new data from different laborites and hospitals via online systems; the next step is an analysis necessary to organize data.

Kidney chronic disease is a dangerous disease. Different methods or techniques are implemented in this field to diagnose this disease at an early stage [4]. Every field of science is doing hard work to overcome it. Computer science is also putting his part in it. Computer science is a vast field, so we select machine learning to detect chronic kidney disease early by using a different methodology [7-8].

Materials and Methods

The implementation of different methods and techniques in image preprocessing to extract limited features and the best algorithm is applied to get the best results. In this chapter, results are obtained and explained using the proposed methodology with the help of CVIPtools version 5.6e and WEKA [7]. These are the best result generation and feature selection software in image processing, mainly in medical science. First, the data set consists of 100 patients' CT scan images in two groups: standard patient CT scan images and patients with chronic kidney disease. Both categories have 50, 50 images. A complete 100 CT scan images are collected in a dataset. In the next step, these images are used in CVIPtools after cropping and resizing the web's small size in an image editor [7]. Cropped images are converted into grayscale levels in utilities. All 100 images are cropped and resized equally, then one by one, changed into color to gray, and created a circle to select the best location. It is an ROI method to select specific parts. Each image has created six circles with (512*512) width, and height column and row are started with (128*128) and radius wand blur radius with (32*32). Each image created six ROIs with the same width, height, and radius, and blur radius the row and column change in each image with the addition of 16 values, so a total of 600 ROIs are created from 100 images. This ROI is saved by selecting binary, histogram, and texture features [8]. These are a total of 21 features. After creating a text file, this file is converted into a CSV file, and then an arff file is created from WEKA software to generate other results.

Results and Discussion

For classification use, WEKA software version 3.8.2, data preprocessing and feature optimization techniques are applied to the dataset, then applied different classification algorithms, namely "Random Forest, MLP, ISO, LogitBoost, BayesNet, and J48" and select the best algorithm of this research. In this chapter, samples of training and testing datasets are given below. Also, different classification results generated by these algorithms and their comparison are given.

In this work, different machine learning classifiers implements on abnormalities of the kidney dataset. This comparative analysis is performed, and other performance calculating parameters learned classification accuracy results.

Random Forest Classifier

- Time taken to build the model: 0.38 seconds
- Test mode: Cross-validation 10

Table 1: Random Forest Classifier Summary

Total Number of Instances	300	
Correctly Classified Instances	289	**96.3333 %**
Incorrectly Classified Instances	11	**3.6667 %**
Kappa statistic	0.9267	
Mean absolute error	0.0817	
Root mean squared error	0.2015	
Relative absolute error	16.3321 %	
Root relative squared error	40.2909 %	

Table 2: Random Forest Classifier Detailed Accuracy

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	Class
0.953	0.027	0.973	0.953	0.963	0.927	0.970	**Normal**
0.973	0.047	0.954	0.973	0.964	0.927	0.970	**Abnormal**
0.963	**0.037**	**0.964**	**0.963**	**0.963**	**0.927**	**0.970**	**Weighted Avg.**

Table 3: Confusion Matrix result using Random Forest Classifier

Classified as	A	B
A = Normal	143	7
B = Abnormal	4	146

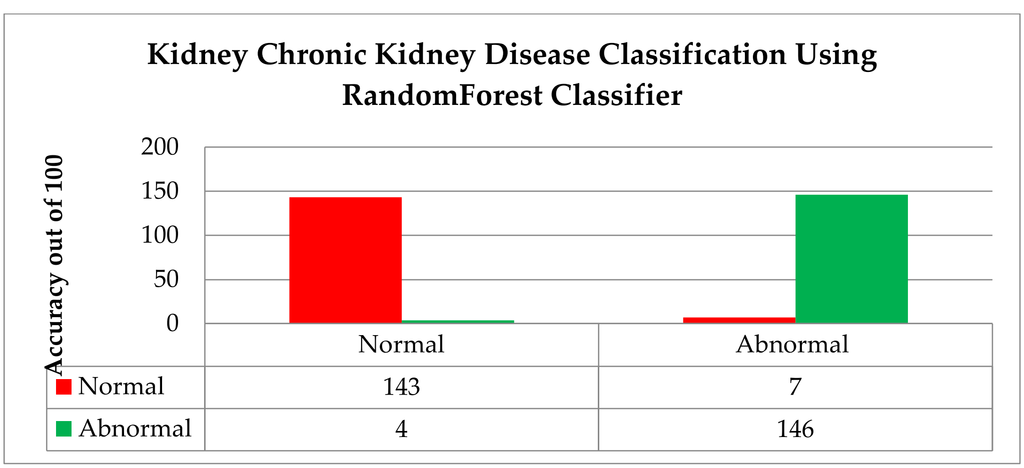


Figure 1: Accuracy of Dataset using Random Forest Classifier

Finally, the overall accuracy of chronic kidney disease has been acquired using the Random Forest classifier 96.33%. The comparison graph of classification accuracy among different machine learning classifiers is shown in Figure 2.

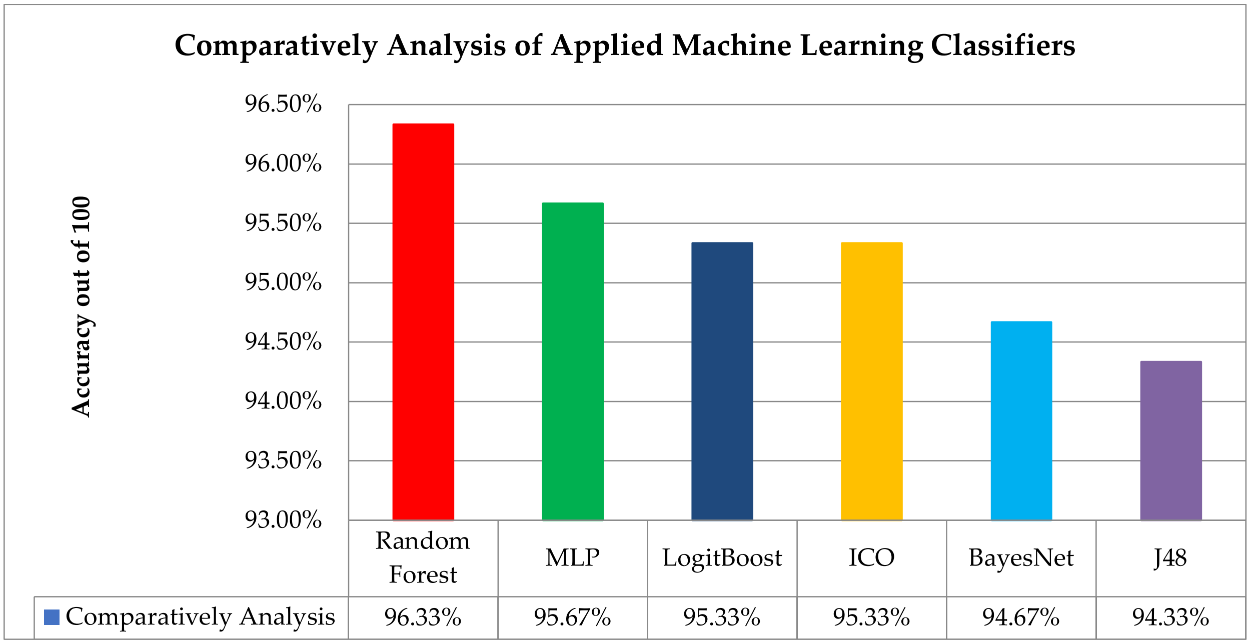


Figure 2: Classification Accuracy Graph Among Different Machine Learning Classifiers

Conclusions

Regardless of some challenges in "chronic kidney disease identification, using machine learning" is the most crucial field of machine learning. Some optimized algorithm of feature extraction extracts normal and abnormal features. The percentage split and cross-validation methods of machine learning are applied to these optimized features. After the training and testing of the processed data, we applied some best classifiers to differentiate the normal and abnormal readings of the dataset. We achieve a maximum of 96.33 % results with Random Forest Tree. Some optimized methods of machine learning and machine vision can improve the accuracy of the results regarding differentiating the abnormal and normal areas in 2D CT-Scans.

References

[1]. Banks, E., Joshy, G., Korda, R. J., Stavreski, B., Soga, K., Egger, S., ... & Lopez, A. D. (2019). Tobacco smoking and risk of 36 cardiovascular disease subtypes: fatal and non-fatal outcomes in a large prospective Australian study. *BMC medicine*, *17*(1), 1-18.

[2]. Faronbi, J. O., Faronbi, G. O., Ayamolowo, S. J., & Olaogun, A. A. (2019). Caring for the seniors with chronic illness: The lived experience of caregivers of older adults. *Archives of gerontology and geriatrics*, *82*, 8-14.

[3]. Flaherty, K. T., Gray, R. J., Chen, A. P., Li, S., McShane, L. M., Patton, D., ... & Conley, B. A. (2020). Molecular landscape and actionable alterations in a genomically guided cancer clinical trial: National Cancer Institute Molecular Analysis for Therapy Choice (NCI-MATCH). *Journal of Clinical Oncology*, *38*(33), 3883.

[4]. Ma, F., Sun, T., Liu, L., & Jing, H. (2020). Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network. *Future Generation Computer Systems*, *111*, 17-26.

[5]. Chen, G., Ding, C., Li, Y., Hu, X., Li, X., Ren, L., ... & Xue, W. (2020). Prediction of chronic kidney disease using adaptive hybridized deep convolutional neural network on the internet of medical things platform. *IEEE Access*, *8*, 100497-100508.

[6]. Canfell, K., Kim, J. J., Brisson, M., Keane, A., Simms, K. T., Caruana, M., ... & Hutubessy, R. (2020). Mortality impact of achieving WHO cervical cancer elimination targets: a comparative modelling analysis in 78 low-income and lower-middle-income countries. *The Lancet*, *395*(10224), 591-603.

[7]. Naeem, S., Ali, A., Qadri, S., Khan Mashwani, W., Tairan, N., Shah, H., ... & Anam, S. (2020). Machine-learning based hybrid-feature analysis for liver cancer classification using fused (MR and CT) images. *Applied Sciences*, *10*(9), 3134.

[8]. Ali, A., Mashwani, W. K., Naeem, S., Uddin, M. I., Kumam, W., Kumam, P., ... & Chesneau, C. (2021). COVID-19 infected lung computed tomography segmentation and supervised classification approach.