Feature Extraction Techniques for Chronic Kidney Disease Identification

Suresh Babu Satukumati[1*], Shivaprasad Satla[2], Raghu Kogila[3]

[1] Department of Information Technology, Vignan's Foundation for Science, Technology and Research (Deemed to be University), Guntur 522017, Andhra Pradesh, India

[2] Department of Computer Science and Engineering, Vignan's Foundation for Science, Technology and Research (Deemed to be University), Guntur 522017, Andhra Pradesh, India

[3] Department of Computer Science and Engineering, VITAE, Hyderabad, India

Corresponding Author Email: suresh.satukumati@gmail.com

ABSTRACT

Chronic Kidney Disease (CKD) is one of the dangerous diseases around the world. Early recognition and appropriate administration are requested for enlarging survivability. According to the UCI informational index, there are 24 qualities for anticipating CKD or non-CKD. At any rate there are 16 qualities need obsessive examinations including more assets, cash, time, and vulnerabilities. The goal of this work is to investigate whether we can anticipate CKD or non-CKD with sensible precision utilizing less number of features. An Intellectual framework advancement approach has been utilized in this investigation. Essential feature determination system to find reduced features that clarify the informational index is introduced. Two insightful paired grouping methods have been received for the legitimacy of the reduced list of capabilities. As recommended from our outcomes, we may more focus on those decreased features for recognizing CKD and along these lines lessens vulnerability, spares time, and reduced costs. The proposed technique uses less features for CKD prediction.

1. INTRODUCTION

Chronic Kidney Disease is a quiet condition. Signs and side effects of CKD, if present, are commonly not explicit in nature and dissimilar to a few other endless infections, (for example, congestive heart fail and constant obstructive lung Disease), they don't uncover a piece of information for finding or seriousness of the condition [1]. The feature extraction process of the kidney is depicted in Figure 1.
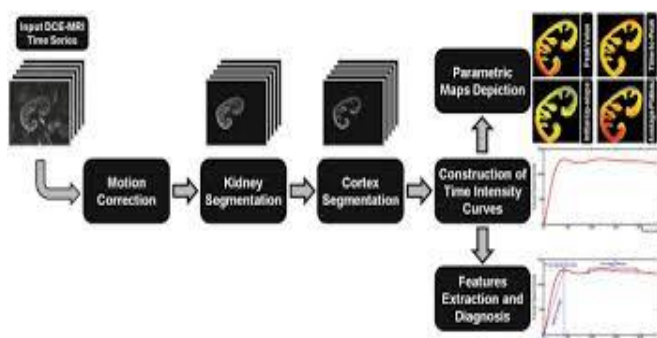

Figure 1. Feature extraction process

Clusters influencing kidney structure and capacity in heterogeneous structure are for the most part named CKD [1]. National Kidney Foundation of America in 2002 made rules for an all the more clear definition and order framework for CKD [2]. CKD is grouped into Stages I– V as indicated by the evaluated Glomerular Filtration Rate (GFR).GFR is assessed having scientific conditions utilizing serum creatinine, age, sex, body estimation, ethnic birthplace, etc. [3] If the ordinary use of a kidney is degraded to a degree, waste can develop to abnormal states in your blood making you feel wiped out [4]. Sadly, this debasement is noted at the later phases of CKD making the issue convoluted in a great part of the cases. Now and then, it is past the point of no return when we counsel a doctor [5].

Consequently, the early location is requested for long haul survivability. It is contended that once in a year or once in 2 years, CKD-related examinations might be finished [25]. General familiarity with the general population must be expanded alongside furnishing less number of neurotic tests with less expense and less time [6].

The phases of constant kidney illness are depicted in Table 1.

Table 1. Phases of CKD

Stage	Clinical features	GFR (mL/min/1.7m^2)
I	Damage with normal or increased GFR	≥ 90
II	Damage with a mild decrease in GFR	60-89
III	Moderate decrease in GFR	30-59
IV	Severe decrease in GFR	15-29
V	Kidney failure	< 15 or dialysis

GFR: Glomerular filtration rate

Presently a-days, Data mining system is joined with

Artificial Intelligence (AI) to separate concealed examples just as for investigation purposes. Information mining is characterized as "a procedure of nontrivial extraction, already obscure and possibly helpful data from the information put away in a database" [7]. Restorative information mining has extraordinary potential like investigating the concealed examples which can be used for clinical conclusion of any illness dataset. Identification of relevant features from the dataset and feature selection and feature extraction plays a very important role in CKD or Non-CKD identification [8].

There are two procedures to perform information mining, to be specific supervised and unsupervised learning. In supervised learning, a preparation set is utilized to learn parameters though in unsupervised adapting no preparation set is utilized. Classification is a directed learning used to find concealed examples from existing restorative information [9].

Clustering is exceptionally basic for treatment of patients. The identification process of CKD involves in utilization of features which is an important task. The Number of features can be reduced for identification process and can be more accurate [10].

2. LITERATURE SURVEY

A cluster based anisotropic distribution sifting system for eliciting the multiplicative commotions in ultrasound pictures which have the huge results. Hu et al [1] proposed a procedure utilizing nearby scale invariant features by processing its distinction of Gaussian, introductions and Hough change for item acknowledgment.

Yu et al. [2] built up a result forecast in immune response inconsistent kidney transplantation by utilizing decision tree and random forest calculations. In their work, the capability of the new transplantation by utilizing DT and RF to anticipate early transplantation with the most elevated execution exactness of 85% is observed.

Wilfrido gomez flores et al. [3] utilized Support Vector Machine grouping calculation for investigation of CKD and they utilized feature selection techniques like wrapper and channel strategies. In their proposed work, they connected distinctive element determination strategies by consolidating property evaluator and internet searcher techniques.

Huang et al. [4] Presented by making arrangement models by utilizing information mining, characterization strategies, for example, NN, ANN, DT and Naïve Bays to anticipate the transitional nterval of kidney illness particularly 3 to 5 phases of kidney ailment. These order models were developed by grouping the chose or decreased arrangement of properties. Subsequent to applying adjusted classifier or quality decreased by the element choice technique, the execution, precision was verified around 85 %

Dehghan et al. [5] utilized decision tree and information mining, arrangement procedures to anticipate CKD for avoidance of the hazard factor of CKD. They actualized CKD dataset on quick mining tools by recovering the dataset to set target job utilizing cross-approval. After order, the execution expectation precision of Naive Bayes was started 86 % and the execution forecast exactness of choice tree begins 91%. In light of the examination result the execution of the decision tree was connected to anticipate CKD.

An edge based anisotropic dissemination channel which focused on both edge safeguarding and commotion decrease is considered. The Adaboost learning incorporating different

capabilities for the characterization of neurotic prostate pictures and to identify small scale calcification in bosom disease pictures. The programmed identification of kidney Diseases utilizing Viola Jones technique consolidated with various features is utilized in advanced mobile phone.

Padmanaban et al. [6] used component determination to extraction rule and foresee CKD by utilizing Naive Bayes classifier and one R attribute selector to keep the CKD. These strategies are with the wrapper subset evaluator with the best first pursuit. After execution on powerful apparatus utilizing wrapper subset evaluators consolidate with the best first web crawler; Naïve Bayes classifier execution found a 97.5 % precision rate.

SVM without feature choice the precision rate was 97.75 %, SVM with the classifier subset join with greedy method the exactness rate was 98 %, SVM with the wrapper subset evaluator consolidate with a best first web search tool the exactness rate was 98.25, SVM with the classifier subset evaluator join with keen feature selection, the precision rate 98.25. Lastly, SVM with the channel subset evaluator combine with best first inquiry the precision was 98.5.

3. PROPOSED METHOD

3.1 Feature extraction for kidney disease detection

Two kinds of features extraction procedures were utilized in this work to be specific Local Binary Pattern (LBP) and Histogram of Oriented Gradients (HOG). Among these two systems, HOG delivered the best location rate on classifier. LBP features [11] are utilized to locate the dimension of the cells taken from Ultra Sound (US) spectroscopy. A relative report [12] with the different sorts of LBP features was investigated. LBP processes the histogram by supplanting neighboring pixel values with parallel qualities for feature extraction [13].

Figure 2. US image of kidney

Figure 3. Detected cyst image

Cluster is the component based group with respect to picture slope, size and posture [14]. This component is generally connected for a few medicinal imaging identification and acknowledgment issues [15]. Split up the picture into square shaped locale called cells. Figure the angle extent and edge introductions for every pixel inside cells [16]. Information pictures are preprocessed by applying Gaussian smoothing and one dimensional fixated cover on both vertical and flat bearing [17].

The casual filter of kidney was made dependent on angle extent and introductions [18]. The standardization of histogram can be determined by gathering nearby cells as a spatial square. Every pixel of the cells has the weighted angles that are moved to comparing precise receptacle [19]. The 9 histogram filter concerning diverse edges from 0 to 160 degrees are made by looking at the estimations of greatness and introduction of the picture subset. At long last the 4x4 subset of a picture with 9 containers of histogram produces 144 element vectors [20]. The slope size is processed by utilizing the condition

$$G = \sqrt{G_x^2 + G_y^2}$$

where, $G_x = I * Gm_x$ and $G_y = I * Gm_x$

The rotation of a picture is determined by utilizing the condition

$$\theta = tan^{-1}(\frac{G_x}{G_y})$$

LBP is perceived as a sort of the visual descriptor for paired surface grouping. It delivers the critical discovery execution on some datasets [21]. Part up the window into cells containing 8X8 pixel, contrast the neighbor pixel and the middle pixel by either clockwise or anticlockwise course, if the inside pixel is more prominent than neighbor pixel, put 0 generally 1 method [22]. The back to back component parallel example considered from the cells is alluded as feature vector [23].

3.2 Connection based feature subset selection and identification

Connection Based Feature Subset Selection (CFS) depends on the accompanying assumption:

A component of a subset is viewed as great which are very related with the class however might be uncorrelated with different features of the class [24].

The element assessment numerical method gives an operational meaning of the above speculation as pursues:

$$r_{fc} = \frac{k\overline{r_{fc}}}{\sqrt{(k + k(k-1)\overline{r_{ff}})}}$$

where, r_{fc} is the connection between's the summed features and the class variable; k is the quantity of features r_{fc} is the normal line of the relationship between's the features and class variable; and r_{ff} is the normal inter connection between's features.

Summarily, an element will be admitted to the subset if its connection with the class is higher than the most elevated relationship among it and anybody of the officially chosen

features.

3.3 Feature extraction from detected kidney diseases

Scale Invariant Feature Transform (SIFT) gives the exceptional example of features, for example, edges, masses, and corners of a picture at specific intrigue point. It gives the gathering of features which has no troubles that are experienced by some other methods [25]. Filter works better regardless of whether progressively number of pictures is caught at a similar area with the diverse positions. Each component vectors are invariant to scaling, interpretation or pivot. The SIFT include is seen by refining strategies. Scale Space Extreme Detection can be considered by getting the extraordinary qualities from the Gaussian convolution. The key point restriction is accomplished by killing poor edges and low differentiation focuses. It expels the huge edge pixels and flow crosswise over edge and little edge neighbors and flow the opposite way. The slope greatness and introduction are determined for each key point. Angle greatness (x, y) can be determined by utilizing the equation

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y)^2 + (L(x,y+1) - L(x,y-1))^2}$$

and the rotation of the image pixels can be

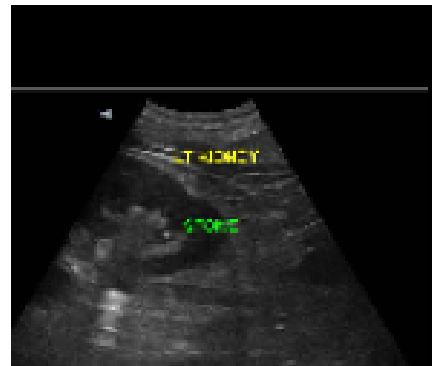$$\theta(x,y) = tan^{-1}\left[\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right]$$

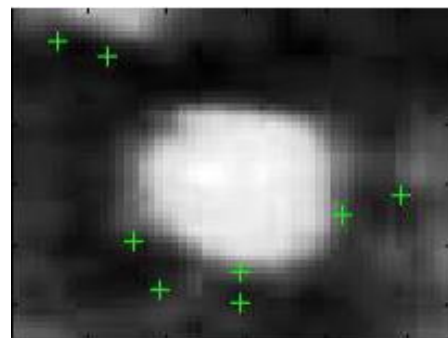**Figure 4.** Ultrasound image with pixel identifications

Figure 5. SIFT key points identified

The proposed calculation is a straightforward probabilistic classifier dependent on applying Bayes' hypothesis with solid autonomy presumptions between the features. Bayes classifiers are profoundly versatile, which need various parameters straight in the quantity of factors in a learning issue.
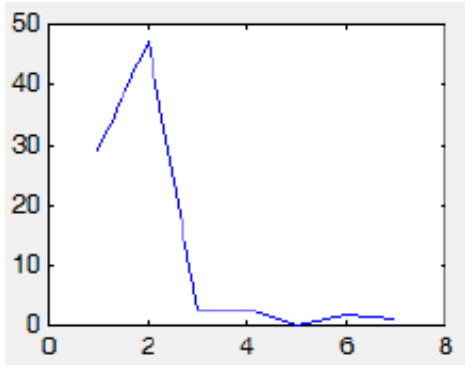
Figure 6. SIFT Features

4. RESULTS

CKD Dataset utilized is taken from UCI repository. It comprises of 11 ostensible traits, 7 numerical qualities and 1 class property. Naïve bayes classifier chooses just 6 properties from 25 all out traits with 76 % of qualities decrease. Best first indexed lists in 12 characteristics from 25 absolute properties with 52 % of qualities decrease.

To foresee the execution of the framework, we figured Cystatin C Ratio (CCR), particularity, affectability, and beneficiary working trademark Area under Curve (AUC) as these are vital parameters to anticipate the execution of the framework without knowing the circulation of information. We registered True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) to additionally figure other execution parameters as talked about underneath:

$$CCR(\%) = \frac{TP+TN}{TP+TN+FP+FN} * 100$$

$$Specificity(\%) = \frac{TN}{TN+FP} * 100$$

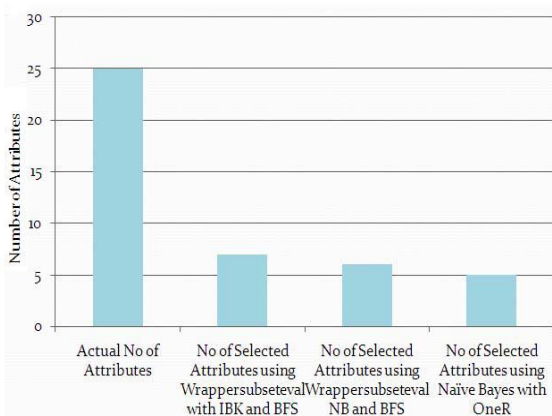$$Spesitivity(\%) = \frac{TP}{TP+FN} * 100$$

Figure 7. Comparison levels of selected features

Different quality evaluators with Attribute Reduction Figure 7, demonstrates the graphical portrayal of precision accomplished by the proposed philosophy which id 97.5 % which is more noteworthy than 12.5 % accomplished by the Naïve Bayes on the first set i.e., 85 %. Additionally we can improve the precision, explicitness and affectability by

utilizing Naïve Bayes with OneR classifier as appeared in Figure 7 which chooses only less features when compared to existing methods.
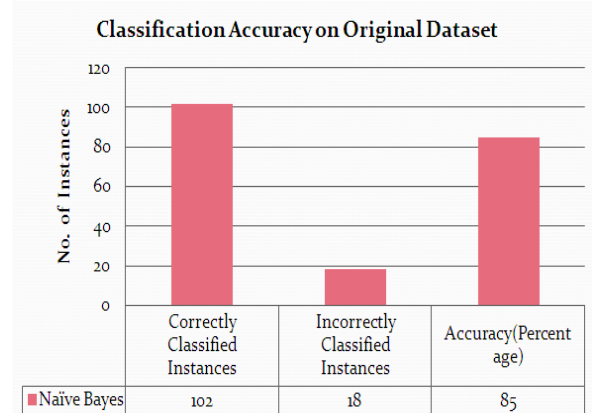

Figure 8. Accuracy of Naïve Bayes by 25 attributes

5. CONCLUSION

The proposed method considers eight features: Explicit gravity, egg whites, serum creatine, hemoglobin, stuffed cell volume, white platelet, red platelet, and hypertension as progressively noteworthy for exploring CKD. Consequently, that these reduced lists of capabilities may be advantageous to investigation when an official conclusion is made by the specialists. Decreased parameters lessen research center expenses and time. In the meantime, reduced features decrease vulnerability in basic leadership. In this work two critical tasks, for example, location and characterization of different kidney Diseases have been tested. Hundred ultrasound B-mode pictures with special size are gathered from different focuses were utilized for this investigation. The quality measures were taken for both location and arrangement. Trial results for the location of different sorts of kidney Diseases utilizing HOG features show better execution.

REFERENCES

[1] Hu ZL, Tang JS. (2016). Driven anisotropic diffusion for speckle reduction in ultrasound images. 2016 IEEE International Conference on Image Processing (ICIP), pp. 2325-2329. https://doi.org/10.1109/ICIP.2016.7532774

[2] Yu JH, Tan, JL, Wang YY. (2010). Ultrasound speckle reduction by a SUSAN-controlled anisotropic diffusion method. Pattern Recognition 43: 3083-3092. https://doi.org/10.1016/j.patcog.2010.04.006

[3] Flores WG, de Albuquerque Pereira WC, Infantosi AFC. (2014). Breast ultrasound despeckling using anisotropic diffusion guided by texture descriptors. Ultrasound in Medicine & Biology 40(11): 2609-2621. https://doi.org/10.1016/j.ultrasmedbio.2014.06.005

[4] Huang CH, Kalaw EM. (2016). Automated classification for pathological prostate images using Adaboost-based ensemble learning. IEEE Symposium Series on Computational Intelligence (SSCI), 1-4. https://doi.org/10.1109/SSCI.2016.7849887

[5] Dehghan F, Abrishami-Moghaddam H, Giti M. (2008). Automatic detection of clustered microcalcifications in

digital mammograms: Study on applying adaboost with SVM-based component classifiers. IEEE 30th Annual International Conference Engineering in Medicine and Biology Society, pp. 4789-4792. https://doi.org/10.1109/IEMBS.2008.4650284

[6] Padmanaban KRA, Parthiban G. (2016). Applying machine learning techniques for predicting the risk of chronic kidney disease. Indian Journal of Science and Technology 9(29). https://doi.org/10.1504/IJIT.2018.090859

[7] Perveen S, Shahbaz M, Guergachi A, Keshavjee K. (2016). Performance analysis of data mining classification techniques to predict diabetes. Procedia Comput Sci 82 (Supplement C), 115-121. https://doi.org/10.1016/j.procs.2016.04.016

[8] Dulhare UN, Ayesha M. (2016). Extraction of action rules for chronic kidney disease using Naive bayes classifier. 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), pp. 1-5. https://doi.org/10.1109/ICCIC.2016.7919649

[9] Borisagar N, Barad D, Raval P. (2017). Chronic kidney disease prediction using back propagation neural network algorithm. Proceedings of International Conference on Communication and Networks Singapore, pp. 295-303.

[10] Pritom AI, Munshi MAR, Sabab SA, Shihab S. (2016). Predicting breast cancer recurrence using effective classification and feature selection technique. 2016 19th International Conference on Computer and Information Technology (ICCIT), pp. 310-314. https://doi.org/10.1109/ICCITECHN.2016.7860215

[11] Mishra S, Chaudhury P, Mishra BK, Tripathy HK. (2016). An implementation of feature ranking using machine learning techniques for diabetes disease prediction. ICTCS '16 Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies 42: 1-3. https://doi.org/10.1145/2905055.2905100

[12] Zufferey D, Hofer T, Hennebert J, Schumacher M, Ingold R, Bromuri S. (2015). Performance comparison of multi-label learning algorithms on clinical data for chronic diseases. Computers in Biology and Medicine 65: 34-43. https://doi.org/10.1016/j.compbiomed.2015.07.017

[13] Bashir S, Qamar U, Khan FH, Javed MY. (2014). MV5: A clinical decision support framework for heart disease prediction using majority vote-based classifier ensemble. Abrabian Journal for Science and Engineering 39(11): 7771–7783.

[14] Baitharu TR, Pani SK. (2016). Analysis of data mining techniques for healthcare decision support system using liver disorder dataset- sciencedirect. Procedia Computer Science 85: 862-870. https://doi.org/10.1016/j.procs.2016.05.276

[15] Sedighi Z, Ebrahimpour-Komleh H, Mousavirad SJ. (2015). Featue selection effects on kidney desease analysis. 2015 International Congress on Technology, Communication and Knowledge (ICTCK), pp. 455-459. https://doi.org/10.1109/ICTCK.2015.7582712

[16] Filimon DM, Albu A. (2014). Skin diseases diagnosis using artificial neural networks. 2014 IEEE 9th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI). https://doi.org/10.1109/SACI.2014.6840059

[17] Ahmad F, Isa NAM, Hussain Z, Osman MK, Sulaiman SN. (2015). A GA-based feature selection and parameter optimization of an ANN in diagnosing breast cancer. Pattern Analysis and Applications Pattern Anal Appl 18(4): 861-870. https://doi.org/10.1504/IJNDC.2017.083642

[18] Ramya S, Radha N. (2016). Diagnosis of chronic kidney disease using machine learning algorithms. International Journal of Innovative Research in Computer and Communication Engineering 4(1): 813-820.

[19] Mohammed SB, Manoj M. (2016). Fused features classification for the effective prediction of chronic kidney disease. International Journal for Innovative Research in Science & Technology 2(10): 44-48. https://doi.org/ 10.1504/IJACT.2012.045589

[20] Ladha L, Deepa T. (2011). Feature selection methods and algorithms. International Journal on Computer Science and Engineering 3(5): 1787-1797.

[21] Tang JL, Alelyani S, Liu H. (2014). Feature selection for classification: A review.

[22] Saeys Y, Inza I, Larranaga P. (2007). A review of feature selection techniques in bioinformatics. Bioinformatics 23(19): 2507-2517. https://doi.org/10.1093/bioinformatics/btm344

[23] Lesley A, Stevens, Zhang YP, MS, Schmid CH. (2008). Evaluating the performance of GFR estimating equations. Journal of Nephrology 21(6): 797-807. https://doi.org/10.1504/IJIDS.2016.075789

[24] Wu XD, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Yu PS, Zhou ZH, Steinbach M, Hand DJ, Steinberg D. (2007). Top 10 algorithms in data mining. Knowledge and Information Systems 14: 1-37. https://doi.org/10.1007/s10115-007-0114-2

[25] Vejendla LN, Gopi AP. (2017). Visual cryptography for gray scale images with enhanced security mechanisms. Traitement du Signal 34(3-4): 197-208. https://doi.org/10.3166/ts.34.197-208