Relative Merits of Data Mining Algorithms of Chronic Kidney Diseases

Harsha Herle[1], Dr. Padmaja K V[2]

Electronics and Instrumentation Engineering, Department
RV College of Engineering, Affiliated to VTU, Bengaluru, India

Abstract—**Early prediction of Chronic Kidney Disease in human subjects is considered to be a critical factor for diagnosis and treatment. The use of data mining algorithms to reveal the hidden information from clinical and laboratory samples helps physician in early diagnosis, thus contributing towards increase in accuracy, prediction and detection of Chronic Kidney Disease. The experimental results obtained from this work, with subjected to optimal data mining algorithms for better classification and prediction, of Chronic Kidney Disease. The result of applying relevant algorithms, like K-Nearest Neighbors, Support Vector Machine, Multi Layer Perceptron, Random Forest, are studied for both clinical and laboratory samples. Our findings show that K - Nearest Neighbour algorithm provides the best classification for clinical data and, similarly, Random Forest for laboratory samples, when compared with the performance parameters like, precision, accuracy, recall and F1 Score of other data mining analysis techniques.**

Keywords—*Ultrasound images; support vector machine (SVM) k-nearest algorithm (K-NN); multilayer perceptron algorithm (MLP); random forest (RF); clinical data*

I. INTRODUCTION

In recent years more than two million people across the globe suffer from Chronic Kidney Disease (CKD) like Kidney stone, kidney transplant, blockage of urine, congenital anomalies, cyst, bacterial infection, dialysis or cancerous cells, to stay alive, of which at least only 10% of the patients need treatment to live with high health care costs [1]. Due to the increase in cost, only 2 million people are capable of receiving treatment for CKD, representing 12% of the global population [2-3]. Within developed countries, only 20% of patients are treated for CKD, and under developed countries, on an average one million die from untreated kidney failure due to financial constraints [9]. CKD can be detected using either X-rays, Ultra Sound (US), Computer Tomography (CT) and MRI or laboratory samples for medication .This work focuses on applying data mining techniques for kidney stone detection using Ultra Sound (US) technique and laboratory samples collected from database, for further treatment by medical doctors.

In detecting the CKD, the laboratory samples obtained from standard database UCI machine learning repository is subjected to data cleaning or preprocessing of the data samples and for further classification, the labels are converted to numbers. The database contains 25 attributes of which serum creatinine, blood ureas, hemoglobin, hypertension remains important in this work and others are out of the scope for our analysis [3].

Once the CKD is detected, the next step is to focus on detecting the kidney stone using US technique. The US technique has more advantages like non-ionising nature, portability, low cost and also in giving the details of real-time monitoring of patient's vital internal organs. US image is recorded by incisive technique, where a high frequency signal of order greater than 1MHz is penetrated into the human body with the help of a transducer. The US waves reflected from kidney tissues are received by transducer and displayed on a computer screen either in two-dimensional (2D) or three dimensional (3D). The obtained US images contain background information and labels that require crossing out and to enhance the quality of US image. Further, the US wave was subjected to speckle noise -multiplicative type of noise that appear as dark and bright spots resulting in trouble analysis and diagnosis interpretation. In reduction of speckle noise in US images, a method termed as speckle denoising is carried out for analysis, preprocessing and interpolation of US images [5]. In removal of speckle noise from US images many algorithms exists and differ in their basic methodologies. Preprocessing filters like, spatial and wavelet filters have proved its efficiency, in-terms of statistical parameters like increased Signal to noise ratio (SNR), Peak Signal to Noise Ratio (PSNR), decrease in Mean Squared Error (MSE), Mean Absolute Error (MAE).

Moreover in US image, the Region of Interest (RoI) is retained and other portions are removed using segmentation techniques. Segmentation is the logical implementation to find the RoI against the characteristic of the US images, thus its features are expected to find the region where kidney stone may be present. Feature extraction method aims at reducing the input data by finding the features from several input patterns, resulting in an input vector consisting of appropriate image properties, which will be given to data mining techniques for further classification.

In classification phase, the input image is classified into abnormal or normal classes depending on the statistical parameters of features obtained from clinical and laboratory data samples. Generally, classification phase is further categorized into unsupervised and supervised classification, where, in supervised classifier, the algorithm iteratively arrives at predictions on the training data and is validated by the teacher, often coined as learning with teacher. Conversely, in unsupervised learning, algorithms are left to their own devices to determine and present the interesting structure in the data, hence does not necessitate training phase for classification. The supervised learning, further classified into,

artificial/logical, Perceptron based, statistical based, and Support Vector Machines (SVM).

This work focus on statistical analysis of clinical and laboratory data samples, where data mining algorithm, K-Nearest Neighbors algorithm, SVM, Multi Layer Perceptron (MLP) and Random Forest (RF) and are used to evaluate the performance of the classifier against different parameters like sensitivity, precision, Recall and F1 score. The experimental results validated the above statistical parameter in estimating the best machine learning algorithm for CKD. Upon classification of kidney stone US images, a Graphical User Interface (GUI) is developed to assist medical doctors about the presence or absence of Kidney US images.

The preceding section in the paper foresees the literature review carried out to find the problem of interest. The research methodology for the problem defined is well stated. The description of constituent blocks and its mathematical relations are well explained and derived. Finally experimental results demonstrate the effectiveness of best suitable algorithm for kidney US Images with conclusion drawn towards it.

II. RELATED WORK

The medical imaging plays a prominent role in detection and diagnosis of diseases related to human subjects, have wide range of scope, thereby researchers and scientists have contributed significantly over decades [6]. To begin with, the laboratory data samples obtained are subjected to transformation and preprocessed or data cleaned for replacing the missing values using data mean method. The data samples are applied to classifiers to evaluate the performance of the algorithms to detect the presence of CKD in human subjects. Upon detection of CKD, the next step is to find the presence of kidney stone using US images.

The raw US image is obtained from the radiologist, unwanted details like anatomical information, are removed by binary threshold method. The US images are subjected to removal of speckle noise either using statistical or classification of model. The main objective in statistical modeling is to remove the noisy images by obtaining the statistical features from training data and later obtain the parameters of interest. Initially, statistical filters such as Weiner filter [20], adaptive filters in spectral domains, finds applicable in removal of additive noise [7]. In order to address the multiplicative noise, Jain mode is proposed [8] that works by taking the logarithmic of the image is obtained, later multiplicative noise is converted to additive noise, and Weiner filtering is applied. This process is tedious, time consuming and depends on the size of the image. A region based segmentation method is applied to kidney along with Gabor filter, resulting in reduction of speckle noise and smoothen the image signal, along with histogram method to improve the quality of the image [9]. The experimental results demonstrate reduction of speckle noise up-to to 85%, focusing on only few parameters. Thresholding methods like soft thresholding, Visu-Shrink, hard thresholding, Sure-Shrink, Bayes sure shrink and Bayes thresholding for speckle reduction were also applied for kidney US imges [10]. The Visu shrink method is based on wavelet shrinkage and uses over smooth images. The Bayes shrink gives promising results in Mean

squared Error (MSE) over visu shrink, all these methods are based on soft thresholding, with the input value is shrunk to zero by the amount of thresholding. In hard thresholding, the input is retained to same value, if it is greater than threshold, else retained to zero [11]. After preprocessing, the next step is to extract the features from US images followed by applying data mining techniques to detect into normal or abnormal using Graphical User Interface (GUI) developed.

The most prominent work involves detection of absence or presence of kidney stones. The US images were segmented using intensity threshold variation that helps in identifying multiple classes to classify the images as stone, early stone stages and normal, [13]. Other methods in feature extraction can be intensity histogram features and Gray Level Co-Occurrence Matrix (GLCM) features. The kidney US images were classified into four different groups like Normal (NR), Bacterial Infection (BI), Cystic Disease (CD) and kidney stones (KS). Thus create the database for classified kidney US image for further pathological studies [12-13].

In classifying the abnormalities in kidney, statistical methods like GLCM or Run Length Matrix (RLM) are used along with SVM, reaching an accuracy of 85.8% [21] Increased in classification accuracy up-to 98.8% can be reached by applying two level set segmentation methods with Artificial Neural Network (ANN) architecture [14]. Intensity histogram and Harlick features were used as feature extraction for segmented Region of Interest (RoI) Kidney US images A two level of classification methods is proposed, of which in the first method, a lookup table based approach was used to classify the US images into normal and abnormal, followed by second stage, where an SVM with MLP was used to classify the presence of stone or cyst in the kidney with promising experimental results reaching an accuracy up-to 98.14%. SVM was used to classify the Kidney US images for early detection of CKD for classifying the training and testing results and to evaluate the performance of SVM with accuracy of 97.6%. The prominent features were extracted from abnormal kidney US images and classified using SVM algorithm reaching accuracy up-to 83.74% [15]. Likewise, the US images were subjected top adaptive median preprocessing method and segmented using K-Means method. GLCM features were extracted and meta-heuristic SVM classifier is used to classify the US images to detect the renal calculi, and have performed better in noisy images exhibiting detection accuracy of 98.8%.

Further, the other machine algorithms like K-NN is used in classification for normal and cyst Kidney US images, with experimental results predicting up-to 92% for normal images and 85% for Cystic images [16]. In addition to this, several machine algorithms like, Logistic Regression, Elastic Net, Lasso Regression, Ridge Regression, SVM, Random Forest, Neural Network, K-NN Elastic Net were applied for kidney US images, of which K-NN alone gives prediction accuracy of 85% [17]. Extending further, a new decision support system was developed to predict and classify CKD using Artificial Neural Network, K-NN, decision tree, Random Subspace, Linear Discriminate Analysis (LDA), of which K-NN alone gives 78% prediction accuracy [19]. Recently, Hybrid classification algorithms play a prominent role in classification of US kidney images, were subjected to SVM-

KNN together, reaching the prediction accuracy up-to 99.6% [17]. Recent advancements in US imaging have revealed the way to increased interest in removal of the speckle noise from the medical images using different algorithms, without reducing much of the diagnostic information [20].

Data mining techniques are applied for laboratory data set consisting of 361 CKD patients. SVM, MLP, Radial Basis function (RBF), Probabilistic neural network (PNN) were applied to these data samples with PNN emerging as the best algorithm that can be used for physicians for further treatment [4].

Of the literature reviewed in this work, the kidney US images are subjected to preprocessing method like Median, Adaptive median, Weiner filter to remove salt and pepper noise, Neigh SURE shrink method to remove speckle noise. The resultant images were applied with GLCM and histogram method for feature extraction. This current work focus data mining technique like SVM, MLP, RF and KNN for the preprocessed Kidney US image to increase the classifier performance. Further the kidney US Images were subjected to data augmentation methods like rotations of US kidney images to enhance the number of images, thereby contributing to increase in overall classification accuracy. The experimental results of both clinical and laboratory data are compared to estimate the optimal data mining algorithm for early diagnosis.

III. RESEACRH METHOD

Fig. 1 shows the block diagram for estimating optimal data mining technique for CKD, using both clinical and laboratory samples. The laboratory data sample obtained from UCI machine learning repository consists of 400 CKD Indian patients and contains 11 numerical and 14 attributes that are categorized. The obtained datasets are preprocessed or data cleaned. The data sample are then passed to different data mining techniques to find the optimal classifier algorithm with respect to accuracy, precision, recall and F1 score. All the preprocessing methods, classification and detection area are carried out in sklearn machine learning API using python programming environment. Upon detection of CKD using laboratory data sets, the next step is to use the same classifiers to detect the presence or absence of kidney stone using Ultrasound Imaging technique.

The raw image is obtained from the radiologist, and unwanted anatomical information are removed using binary threshold method. The resultant noisy image is then filtered to suppress the Salt and Pepper noise, speckle noise using Median, Adaptive Median, Weiner and Neigh SURE shrink filtering method. The preprocessed image is segmented by threshold method and morphologically operated to detect RoI in kidney, and then the statistical features are extracted from normal and abnormal images to classify the kidney abnormalities. If the abnormalities are detected, then the Centroid of the abnormal image is estimated to find the area of kidney abnormal region. The detected abnormalities will provide information for medical doctors, to take the medication to next level. All the preprocessing methods, classification and detection of centroid area for kidney stone detection are carried out in image acquisition tool box in

MATLAB® 2018a version simulation environment for image resolution of 256×256 using high end computers.

A. Preprocessing

In preprocessing the laboratory samples, data sets are subjected to interpolation, transformation and scaling, where interpolation method includes mean, median or most frequent values. Data transformation, method converts label to number and scaling is a process to normalize the values in the data set. Likewise, for Kidney US images, the acquired image undergo different preprocessing stages to increase the quality of images [39]. The acquired images are prone to speckle noise, salt and pepper noise, which requires to be filtered out. In this work, Neigh SURE shrink, median, adaptive median, Weiner filter are used to remove the both the noise and un-sharp masking for sharpening. Entropy based segmentation is used for finding the RoI and morphological operations like erosion and dilation are also used for finding the final segmented image.

1) Median filters: Median filters are non-adaptive filters that adjust the coefficients based on data within a rigid moving window, and summons between the quality of speckle noise suppression and the potential to preserving image details. They are linear statistical filter which works with the principle of replacing the current pixel value in US images into the median value in a neighborhood. [19-20].

2) Adaptive median filers: This filter works with the three step methodology of which the primary step is to detect the noise, in US images, secondly, adaptively estimate the window size based on the number of noise pixels within the window. Finally, determine the weight of each non-noise point in filtering window and filter off noise points by means of weighted median filtering algorithm. This method is advantageous compared to other preprocessing filters, as it preserves the edges of its high frequency parts of an image. [25].
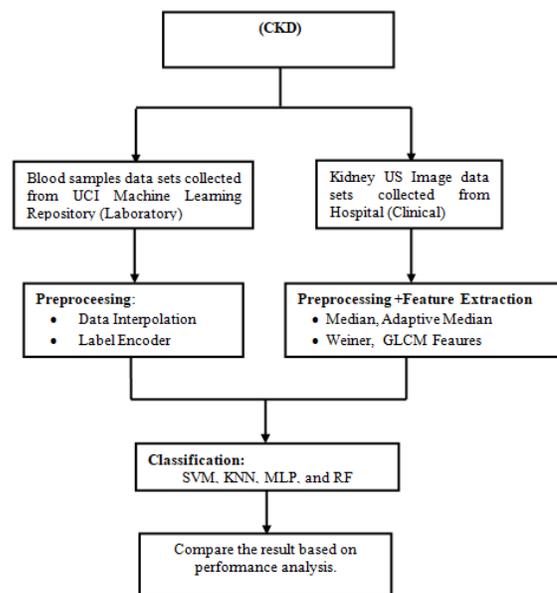

Fig. 1. Block Diagram for Estimating the Optimal Data Mining Algorithm for CKD.

3) Weiner filters: Another common preprocessing filter that finds application in US image is Weiner filter that works on the principle of blurring and removal of additive noise simultaneously by performing best possible substitution between inverse filter and noise smoothing [18] and is optimal in terms reduction in Mean Square Error [22].

4) Neigh SURE shrink filters: Neigh SURE shrink is the adaptive filter based method works on the principle of thresholding technique, combining both soft thresholding and hard thresholding, and only deals with binary values. In this method, a fixed threshold value is considered based on the window size, $[W_s]$ and variance value, is calculated using the equation 1,

$$\sigma^2 = \frac{\text{median}(|w_s|)}{0.6745} \ (w_s \in \text{subband LL}) \qquad (1)$$

The smooth or low frequency components are approximated from father wavelet and high frequency components are resembled from mother wavelet. A 3x3 sized window is moved all over the image to be filtered, where the surrounding pixel value is averaged over the central pixel. Mean and variance is applied over the decomposed image to reduce the speckle noise from US image reduction in speckle noise is based on the thresholding value, when the pixel value crosses the threshold value then that pixel value becomes 1 or 255 and if the pixel value is below the threshold value then that pixel value becomes 0. The Neigh Shrink uses a suboptimal universal threshold and identical window size in all wavelet subbands, whereas the improved version of it determines an optimal threshold and neighboring window size for every subband by the Stein's unbiased risk estimates [23] as given in equation 2.

$$(T^s, k^s) = \arg \min {}_{T,k} \text{SURE}(w_s, \rho, k) \qquad (2)$$

Where ρ is the threshold, k is the window size and s denotes the sub band [24].

B. Feature Extraction

Feature extraction primarily focuses on reducing the input data by determining the distinguishable features from numerous input data samples. The output of feature extraction stage results in an input vector consists of significant image properties, which will be fed into classifier to classify normal and kidney stone US images. In medical imaging analysis, texture is an important feature that provides the spatial distribution of pixels gray level in a region. Textural characteristics of particular image is extracted, to perform the identification of object regions, using many diverse algorithms like fractal based methods, Markov random field and Gabor filter. Harlick features often coined as Gray level co-occurrence matrix (GLCM) is a statistical feature extraction method, represented in the form, $N_g \ X \ N_g$, where Ng is number of gray levels, Within MATLAB, GLCM is represented in the form of a matrix where number of rows and columns are equal to number of gray levels, G in the image. GLCM gives spatial relationships between pixels P(i,j) [26].

Of various features, this work focuses on homogeneity, contrast, correlation, and energy. Another prominent feature extraction method, is Gray Level Run Length Matrix (GLRLM), can be used to extract the higher order statistical features in a US kidney images. Unlike GLCM, the GLRLM is a two dimensional matrix of *Ng X R* element in which each element P (k, l|θ) gives the total occurrence of runs having length k of gray level L in a given direction, with R repenting the longest run. All the seven statistical features namely, Short Runs Emphasis, Long Runs Emphasis, Gray Level Non Uniformity, Run Length No uniformity, Run Percentage, Low Gray level Runs Emphasis, and High Gray level Runs Emphasis are used in this work. [21].

C. Data Mining / Classifer

The principle behind the classification stage is to categorize the input image into normal or abnormal class depending on the features extracted, based on statistical parameters. The data mining techniques like SVM, KNN, MLP and RF are used in the present work for classification of clinical and laboratory data samples.

1) Support Vector Machine (SVM): SVM classification is a non probabilistic linear binary classifier that analyzes the input data and predicts one of the two classes it belongs to, that are separated by hyper plane, constructed using structural risk minimization principle. The SVM is independent of the dimensionality of the feature space that outperforms as compared to other classifiers with minimum training samples. [27].

2) K- Nearest Neighbour (KNN): Another important classifier used in clinical and laboratory data samples, K-NN classifier that classify the datasets according to majority of its neighbors belongs to. Selection of number of neighbours is elective and user defined. [17].

To select the K, for the data sets, the KNN algorithm is executed several times for different values of K, the optimal value of K is chosen so that number of errors is reduced. Euclidean distance is one of frequently used method for distance measures to find the K-NN and is given by,

$$Euclidean = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \qquad (3)$$

The minimum distance between the training and test samples gives best classification of the test samples.

3) Multilayer Perceptron Algorithm (MLP): Another significant class of classifier algorithm is MLP, consists of an input layer, one or more hidden layers, and the output layer, of which the output in different stage is activated either using linear or non-liner activation function. This algorithm is sub classified into two stages, wherein the feed forward stage, the output is estimated by weightage average of inputs and bias. In the backward stage, the errors are minimized by updating the weights [28].

4) Random Forest (RF): Random forest algorithm works with the principle of creating the decision trees on data samples obtained, accumulate the prediction from each of the samples, and provides the optimal solution through voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result.

Generally RF consists of many decision trees, and the features are randomly selected from the optional features, thus allowing the tree at each node to grow without trimming. RF algorithms reaches high accuracy even for a portion of the data is missing in the data samples [29].

IV. RESULT AND DISCUSSION

During the preprocessing stage for all the 400 laboratory samples, the patient ID is removed from the dataset and missing predictor attributes/variable values are replaced using data interpolation method. For the categorized attributes like Hypertension, Coronary Artery Disease, Appetite, Pedal Edema, Pus Cells, Pus Cells Clumps, Diabetes Mellitus, and Anemia, the digitizing label encoder was also applied. The resultant data sets are further subjected to data mining techniques as mentioned above. Each data attribute contains distinct and unique features that also help in classification of CKD, which was shown in Table I.

TABLE I. ATTRIBUTES DESCRIPTION USED IN THE ANALYSIS (PRED-PREDICTOR, NUM-NUMERICAL, NOM-NOMINAL, TAR-TARGET)

Sl.NO.	Dataset Attributes	Class	Type	Misssing	Distinct	Unique
1	Age	Pred	Num	9	76	16
2	Blood Presure	Pred	Num	12	10	3
3	Specific Gravity	Pred	Nom	47	5	0
4	Albumin	Pred	Nom	46	6	1
5	Sugar	Pred	Nom	49	6	0
6	Red Blood Cells	Pred	Nom	152	2	0
7	Pus Cells	Pred	Nom	65	2	0
8	Pus Cells Clumps	Pred	Nom	4	2	0
9	Bacteria	Pred	Nom	4	2	0
10	Blood Glucose Random	Pred	Num	44	146	65
11	Blood Urea	Pred	Num	19	118	55
12	Serum Creatinine	Pred	Num	17	84	41
13	Sodium	Pred	Num	87	34	7
14	Potassium	Pred	Num	88	40	8
15	Hemoglobin	Pred	Num	52	115	28
16	Packed Cell Volume	Pred	Num	71	42	8
17	White Cell Blood Count	Pred	Num	106	89	31
18	Red Cell Blood Count	Pred	Num	131	45	3
19	Hypertension	Pred	Nom	2	2	0
20	Diabetes Mellitus	Pred	Nom	2	2	0
21	Coronary Artery Disease	Pred	Nom	2	2	0
22	Appetite	Pred	Nom	1	2	0
23	Pedal Edema	Pred	Nom	1	2	0
24	Anemia	Pred	Nom	1	2	0
25	Class	Tar	Nom	0	2	0

The next step is to use the mentioned-preprocessing filters for speckle noise reduction, applying for both normal and kidney stone US images of clinical data samples. The different spatial filters like, Median, Adaptive median, Weiner are used in estimating statistical parameters. Further, for the different noise variance (NV) varying from 0.01 to 0.08, the statistical parameters like Peak Signal to Noise Ratio (PSNR in dB), Signal to Noise Ratio (SNR in dB), Root Mean Square Error (RME) and Mean Absolute Error (MAE) are estimated for both normal and abnormal kidney images.

Further, for the US images, the experimentation begins with removal of background and labels associated with kidney images for both normal and kidney stone US Images, followed by enhancing the contrast and sharpening the quality of image for better performance as shown Fig. 2.

Fig. 2. (a) Normal Kidney Image. (b) Removal of Background and Label of Kidney Images. (c) Contrasted Image and Sharpened Image.

Table II represents statistical analysis of noise variance for different filters used in preprocessing of US images. The NV values from 0.01 to 0.08, for different filters, the statistical parameters like PSNR, SNR, MAE and RME are calculated and only the optimal values are provided in Table II. For Kidney stone US images, the adaptive median filter with NV value of 0.01 gives PSNR as 33.51dB and SNR of 18.01dB as compared to other filters. Likewise, the same method was applied to normal kidney US images and results are tabulated. The presence of kidney stone decreases the PSNR and SNR as compared to normal US images with the variation in RME and MAE.

Fig. 3 shows the preprocessed normal US images, from (a-c), and kidney stone US images from (d-f). It is evident that the speckle noise is reduced relatively better using median filter, but image is blurred and artifacts are introduced, as compared with other filters. In case of Adaptive filter, maximum speckle noise reduction is observed, and the edges features are well preserved. In Weiner filter, image is subjected to over enhancement, resulting in reduction in speckle noise, but posing diagnosis problems. Likewise, the same US images are also subjected to Neigh SURE Shrink wavelet filters with haar, db4, sym-8, and for decomposition levels 2 and 4.

From Fig. 4, for decomposition level-4, Wavelet Neigh SURE shrink along with haar, db-4, sym8, increases the contrast resolution that leads to superior visual quality. This, in turn, enhances the quality of kidney stone boundaries, reducing the speckle noise, increasing the SNR, when compared to decomposition level-2, thus, improving the image quality for further diagnosis. Likewise, comparing the decomposition level 2 and level 4, it is observed that decomposition level 4 improves the PSNR and SNR.

TABLE II. STATISTICAL ANALYSIS OF NV AGAINST DIFFERENT PREPROCESSING FILTERS

Kidney Stone US image					
Filter	Noise variance	PSNR in dB	SNR in dB	RME	MAE
Median	0.01	31.73	16.23	6.60	1.49
Adaptive Median	**0.01**	**33.51**	**18.01**	**5.38**	**1.70**
Weiner	0.04	31.85	16.35	6.51	2.58
Normal Kidney US image					
Median	0.01	30.84	18.50	7.31	3.74
Adaptive Median	0.01	33.70	21.36	5.26	2.76
Weiner	**0.05**	**35.29**	**23.29**	**4.21**	**2.24**

TABLE III. STATISTICAL ANALYSIS OF NV AGAINST DIFFERENT WAVELET FILTERS

Kidney Stone US image					
Neigh SURE shrink	Decomposition Level	PSNR in dB	SNR in dB	RME	MAE
db4	4	41.62	26.12	2.11	1.34
Haar	4	41.33	25.83	2.18	1.37
sym8	**4**	**41.86**	**26.36**	**2.03**	**1.30**
Normal Kidney US image					
db4	4	41.70	29.36	2.09	1.46
Haar	4	40.10	27.75	2.51	1.72
sym8	**4**	**42.21**	**29.87**	**1.97**	**1.39**

Fig. 5 depicts the GUI developed using MATLAB, to detect the presence/absence of the kidney stone in US images Upon GLCM and GLRLM feature extraction, the next step in image processing is to classify the kidney US Images into normal and kidney stone US images. The normal and kidney stone US Images together are used in training and validation process for classification. A data mining/Classifier performance measure is based on accuracy, precision, recall and F-1 Score, i.e. number of sample that can be into normal or abnormal classes, hence this work proposes the application of SVM , KNN, MLP and RF classifier for the classification US images with tenfold cross validation to train and validate the classifiers.
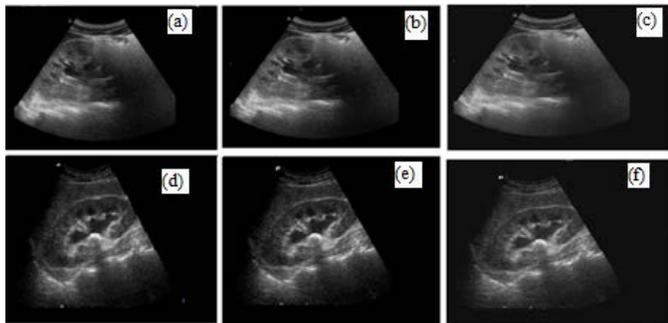

Fig. 3. (a-c) Results of Median Filter, Adaptive Median, Weiner Filter for Normal US Image (d-f) Results of Median Filter, Adaptive Median, Weiner Filter for Kidney Stone US Images.
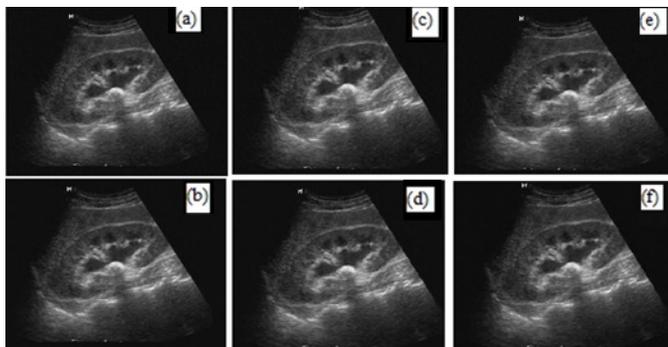

Fig. 4. Preprocessing Filters of Kidney Stone US Image, (a-b) Sys8 Level 2 and 4 Decomposition, (c-d) db4 Level 2 and 4 Decomposition, (e-f) Haar Level 2 and 4 Decomposition.

In Table III, it is observed that Neigh SURE shrink with sym8 for decomposition level 4 of kidney stone images, PSNR is found to be 41.82dB and SNR of 26.36dB, along with reduction in RME and MAE respectively. The presence of kidney stone in US images reduces PSNR and SNR values, compared to normal kidney images. Comparing Tables II and III, the increase in PSNR and SNR, reduction of RME and MAE can be observed. In Conclusion, sym8 Neigh SURE shrink filter is a far superior filter that can be employed in US Image preprocessing method.
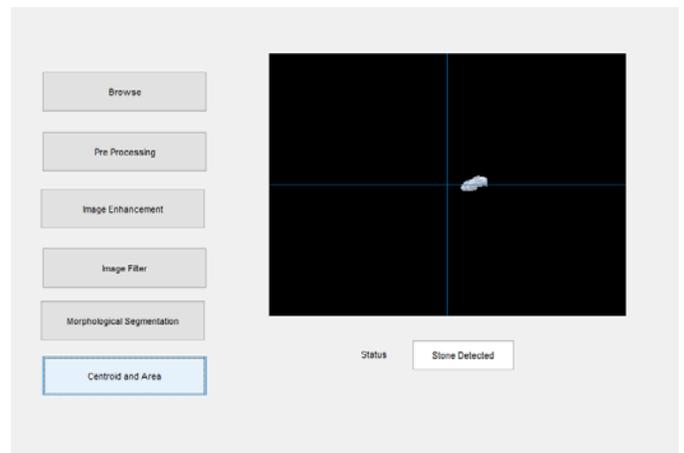

Fig. 5. RoI along with Graphical User Interface (GUI) to Detect the Presence of Kidney Stone.

True positive, True Negative, False Positive and False Negative are the confusion matrix features that are used for measuring the accuracy, precision, Recall, and F-1 Score of the classifier system. All these parameters are characterized by below relation.

$$\text{True Positive (TP)} = \frac{\text{Number of Image with kidney stone}}{\text{Total Number of kidney Images}}$$

$$\text{True Negative(TN)} = \frac{\text{Number of Images without kidney stone}}{\text{Total number of kidney Images}}$$

$$False\ Positive(FP) = \frac{Number\ of\ Images\ falsely\ detected\ as\ kidney\ stone}{Total\ Number\ of\ kidney\ Images}$$

$$False\ Negative(FN) = \frac{Number\ of\ Images\ having\ kidney\ stone\ but\ not\ detected}{Total\ Number\ of\ kidney\ Images}$$

$$False\ Negative(FN) = \frac{Number\ of\ Images\ having\ kidney\ stone\ but\ not\ detected}{Total\ Number\ of\ kidney\ Images}$$

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F1\ Score = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

Table IV depicts the classifier category of clinical and laboratory samples using different data mining techniques. From Table IV it is observed that, the KNN algorithm yields better results of 100% precision, accuracy against other classifier for clinical sample or kidney US images. Subsequently increase in accuracy precision, Recall and F-1 Score are also noticed in below table. Likewise, for laboratory samples, RF algorithm reached maximum accuracy of 100% along with other performance parameters against different data mining techniques.

TABLE IV.   PERFORMANCE EVALUATION FOR CLINICAL AND LABORATORY DATA SAMPLES

Performance Evaluation (%)				
Clinical data sets- US images				
Classifier	**Precision**	**Recall**	**F1 Score**	**Accuracy**
Support Vector Machine	0.77	0.62	0.58	0.62
K Nearest Neighbour	**1.0**	**1.0**	**1.0**	**1.0**
Multi Layer Perceptron	0.58	0.57	0.56	0.57
Random Forest	0.88	0.87	0.87	0.87
Laboratory data sets- Blood sample				
Support Vector Machine	0.61	0.61	0.57	0.61
K Nearest Neighbour	0.88	0.88	0.88	0.88
Multi Layer Perceptron	0.83	0.71	0.70	0.71
Random Forest	**1.0**	**1.0**	**1.0**	**1.0**

Furthermore, the experimental results obtained from the work are compared with the results obtained from similar works and are tabulated as depicted in Table V. By optimal selection of nose variance value from the results obtained, it is observed that the machine learning algorithm KNN when applied to US images reaches a maximum accuracy, whereas the Random Forest algorithm yields an accuracy of 87%. Extending further, for the laboratory samples, the experimental results obtained for Random forest with that of enhanced decision tree algorithm, with KNN reaching an accuracy of 88%. Of all the comparative analysis carried out, it is evident from the experimental results that KNN best suits for detecting the Kidney stone using US technique.

TABLE V.   COMPARATIVE ANALYSIS OF DATA MINING METHODS FOR CKD

Ref.No.	Kidney category	Feature Extraction	Classifier	Accuracy (%)
A. Clinical Data samples				
[30]	Normal, abnormal	PCA	SVM, KNN	84, 89
[21]	Normal, abnormal	Gradient features, GLCM	SVM	85.8±3.1
[31]	Normal, abnormal	Statistical features GLCM	SVM	85.7
[32]	Normal, abnormal	GLCM	ANN	87.5
[33]	Normal, abnormal	Gobar features	ANN	87.06
[34]	Normal, abnormal	SVD Features	SVM	90
[13]	Normal, abnormal, bacterial infection, cystic diseases	GLCM and intensity histogram	ANN	-
[35]	Normal, abnormal	Statistical wavelet based features+ PCA	ANN	97
[36]	Normal, abnormal	GLCM+GLRLM+ Differential Evolution (DE)	SVM	86.3
[37]	Normal, abnormal	Statistical features	SVM	84
[38]	Normal, abnormal	Gabor wavelet features	ANN	-
[39]	normal, kidney stone, cystic, kidney tumor	GLCM+PCA	ANN	77.8
[40]	Normal, abnormal	GLCM+PCA	SVM,KNN	84,89
[41]	Normal, abnormal	K-Means and GLCM features	Meta-Heuristic SVM	98.8
[42]	Normal, abnormal	wavelet energy features	ANN	96.8
This work	**Normal, abnormal**	**GLCM+GLRLM**	**KNN, RF**	**100,87.0**
B. Laboratory data				
[5]	Normal, abnormal	Physiological and Clinical Attributes	PNN, RBF, SVM,MLP	96.7,87 60.7, 51.5
[43]	Normal, abnormal	Physiological and Clinical Attributes	Ant Colony based Optimization (D-ACO)	95
[44]	Normal, abnormal	Physiological and Clinical Attributes	Decision Tress.	99.5
[45]	Normal, abnormal	Physiological and Clinical Attributes	Enhanced Decision Tree	100
[46]	Normal, abnormal	Physiological and Clinical Attributes	SVM,KNN	97.75, 98.5
This work	**Normal, abnormal**	Physiological and Clinical Attributes	**RF,KNN**	**100,88.0**

V. Conclusion

The main focus of this work is to find the optimal data mining algorithms for both clinical and laboratory data sets. This work also characterize the NV parameter varying from 0.01 to 0.08 against the statistical parameters like SNR, PSNR, RME and MAE for different preprocessing filters like spatial and wavelet filters for US images, and find the best filter can be used for preprocessing the US Images. From the experimental results it is evident that, sym8 decomposition level 4, provide increased SNR (29.87dB), PSNR (42.21dB), with reduction in RME (1.97) and MAE (1.3921) as compared other filtering methods. Further, morphological operations like erosion and dilation were applied to segment the filtered US image. By applying entropy based segmentation and morphological operations, feature extraction, RoI and exact area of kidney stone can be located for kidney stone US image. In classifying clinical and laboratory data sets different data mining techniques such as SVM, KNN, MLP and RF are used. The RF and KNN reaches good accuracy up-to 100% for laboratory and clinical data sets against other techniques discussed above.

The notable limitation of this works is direct comparison of experimental result obtained with result obtained from similar work is beyond scope of this work, as kidney US Images obtained vary between hospitals. To author knowledge this is the first work to use NV from 0.01 to 0.08 against statistical parameters. Further increase in noise value may not directly improve the PSNR, SNR or decreases in RME and MAE.

This work recommends the use of KNN for kidney US Images and RF for laboratory samples that can be used for physicians in order to eliminate diagnostic and treatment errors. To assist the medical doctor for further treatment, a GUI was developed, that helps in detection of kidney stone and its area with minimum effort.

References

[1] William G Couser , Giuseppe Remuzzi, Shanthi Mendis, Marcello Tonelli, "The contribution of chronic kidney disease to the global burden of major non communicable diseases", *Kidney Int, 2011*, , vol. 80, no. 12, December 2011, DOI: 10.1038/ki.2011.368.

[2] Vivekanand Jha, Guillermo Garcia-Garcia, Kunitoshi Iseki, Zuo Li, Saraladevi Naicker, Brett Plattner, Rajiv Saran, Angela Yee-Moon Wang, Chih-Wei Yang "Chronic kidney disease: global dimension and perspectives", *Lancet*, Vol 382, pp.260-273, July 2013.

[3] Andrew S Levey , Josef Coresh, Ethan Balk, Annamaria T Kausz, Adeera Levin, Michael W Steffes, Ronald J Hogg, Ronald D Perrone, Joseph Lau, Garabed Eknoyan, National Kidney Foundation, "National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification", Annals of internal medicine, Vol.139, No.2,pp. 137-147, July 2003.

[4] Dua, D. and Graff, C. (2019). "UCI Machine Learning Repository" [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[5] El-Houssainy A.Rady, Ayman S.Anwar, "Prediction of kidney disease stages using data mining algorithms", Elsevier Journal, Informatics in Medicine Unlocked: Vol. 15,pp.1-7,April 2019.

[6] R.F. Wagner, S.W. Smith, J.M. Sandrik, H. Lopez, "Statistics of Speckle in Ultrasound B-Scans", *IEEE Transactions on Sonics and Ultrasonics* Vol 30, No.3, pp.156-163,May 1983.

[7] Kunio Doi, Ph.D., "Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential", Comput Med Imaging Graph., Vol 31, Issue (4-5), pp.198–211, 2007.

[8] J. Portilla; V. Strela; M.J. Wainwright; E.P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain" IEEE Transactions on Image Processing, Vol 12, Issue 11, pp-1338-1351,2003,

[9] Carlos A. Duarte-Salazar, Andrés Eduardo Castro Ospina, Miguel A. Becerra, Edilson Delgado-Trejos, ," Speckle Noise Reduction in Ultrasound Images for Improving the Metrological Evaluation of Biomedical Applications: An Overview" Vol 8, pp 15983 – 15999,2020.

[10] Tanzila Rahman, Mohammad Shorif Uddin, "Speckle noise reduction and segmentation of kidney regions from ultrasound image", 2013 International Conference on Informatics, Electronics and Vision (ICIEV), DOI:10.1109/ICIEV.2013.6572601.

[11] D.Gnanadurai, V.Sadasivam, "An Efficient Adaptive Thresholding Technique for Wavelet Based Image Denoising", Published 2006, World Academy of Science, Engineering and Technology, International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering, Vol 2, No 8, pp 1703-1708, 2008.

[12] Quan Pan, Lei Zhang, Guanzhong Dai, and Hongcai Zhang, "Two Denoising Methods by Wavelet Transform", IEEE transactions on signal processing, Vol. 47, NO. 12,pp 3401-3406, December 1999.

[13] Hafizah, Wan Mahani and Supriyanto, Eko and Yunus, Jasmy, *"Feature extraction of kidney ultrasound images based on intensity histogram and gray level co-occurrence matrix", In Proceedings* - 6th Asia International Conference on Mathematical Modelling and Computer Simulation, IEEE, New York, USA, pp. 115-120, 2012.

[14] Wan Mahani Hafizah, Eko Supriyanto, "Automatic Region of Interest Generation for Kidney Ultrasound Images", International Journal of Biology and Biomedical Engineering, Vol 6, Issue 1,pp 26-34, Oct 2012.

[15] K. Viswanath R. Gunasundari, "Modified Distance Regularized Level Set Segmentation Based Analysis for Kidney Stone Detection", International Journal of Rough Sets and Data Analysis (IJRSDA),Vol 2, Issue 2, pp 18, 2015.

[16] Yan Chen, Ruming Yi, Patrick Flynn, Shira Broschat, "Aggressive region growing for speckle reduction in ultrasound images", , Pattern Recognition Letters Elsevier, Vlo 24, Issue 4-5, pp 677–691,2003.

[17] Prema T. Akkasaligar, Sunanda Biradar, "Classification of Medical Ultrasound Images of Kidney", IJCA Proceedings on International Conference on Information and Communication Technologies ,2014.

[18] Jing Xiao, Ruifeng Ding, Xiulin Xu, Haochen Guan, Xinhui Feng, Tao Sun, Sibo Zhu, Zhibin Ye, "Comparison and development of machine learning tools in the prediction of chronic kidney disease progression", Journal of Translational Medicine ,Vol 17, pp 119, 2019.

[19] Amit GARG, Vineet KhandelwaL, "Speckle Noise Reduction in Medical Ultrasound Images Using Modelling of Shearlet Coefficients as a Nakagami Prior", Advances in Electrical and Electronic Engineering, Vol 16, No4, pp-538-549, 2018.

[20] Ramandeep Kaur; Akshay Girdhar; Jappreet Kaur, "A New Thresholding Technique for Despeckling of Medical Ultrasound Images", Fourth International Conference on Advances in Computing and Communications IEEE, 2014.

[21] M. B. Subramanya, Vinod Kumar, Shaktidev Mukherjee & Manju Saini ,"SVM-Based CAC System for B-Mode Kidney Ultrasound Images",Journal of Digital Imaging, Elsevier, pp 448-458,Dec 2015.

[22] A.D. Hillery and R.T. Chin, "Iterative Wiener filters for image restoration", IEEE Transactions on Signal Processing, Vol 39, No. 8, pp. 1892 – 1899, 1991.

[23] Zhou qin-wu et.al,(2002) "Denoise and contrast enhancement of ultrasound speckle image based on wavelet", 6th International Conference on Signal Processing, Beijing, China, 2002.

[24] Q.Pan, "Two denoising methods by wavelet transform", IEEE Transactions on Signal Processing, Vol. 47, pp.3401–3406,1999.

[25] T. Loupas, W.N. McDicken, P.L. Allan, "An adaptive weighted median filter for speckle suppression in medical ultrasonic images", IEEE Transactions on Circuits and Systems, Vol. 36, No.1, pp-129-135, January 1989.

[26] Pallavi Vaish, R Bharath, P Rajalakshmi, U. B. Desai, "Smartphone Based Automatic Abnormality Detection of Kidney inUltrasound

Images", IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom),2016.

[27] Chih-Wei Hsu, Chih-Jen Lin "A comparison of methods for multiclass support vector machines", *IEEE transactions on Neural Networks*, Vol. 13, No. 2, pp. 415-425, March 2002.

[28] Vinoth R, Anusha C, SaiPriya D, Sridhar BS, ManiKantaPavan C, Aditya G. "A Novel Approach on Ultrasound Kidney Images Classification by using Micro Deep Neural Network Algorithm", International Journal of pure and applied Mathematics, Vol. 119, No. 12, pp.15047-15058., 2018.

[29] Angshuman Paul, Dipti Prasad Mukherjee, Prasun Das, Abhinandan Gangopadhyay, Appa Rao Chintha, Saurabh Kundu, "Improved Random Forest for Classification", IEEE Transactions on Image Processing ,Vol.27, No.8, pp. 4012 – 4024, Aug. 2018.

[30] J. Verma, M. Nath, P. Tripathi, and K. Saini, "Analysis and identification of kidney stone using kth nearest neighbour (KNN) and support vector machine (SVM) classification techniques," Pattern Recognition and Image Analysis, vol. 27, no. 3, pp. 574–580, 2017.

[31] K. Sharma and J. Virmani, "A decision support system for classification of normal and medical renal disease using ultrasound images: A decision support system for medical renal diseases," International Journal of Ambient Computing and Intelligence, vol. 8, no. 2, pp. 52–69, 2017.

[32] M. Pawar and A. Mulla, "Design and analysis performance of kidney cyst detection from ultrasound images," International Journal of Engineering Development and Research, vol. 5, no. 4, pp. 911–917, 2017.

[33] S. Chaitanya and A. Rajesh kumar, "Classification of kidney images using particle swarm optimization algorithm and artificial neural networks," International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 4, pp. 526–530, 2019.

[34] S. Sudharson and P. Kokil, "Abnormality classification in the kidney ultrasound images using singular value decomposition features," 2019 IEEE Conference on Information and Communication Technology, 2019, pp. 1-5.

[35] Attia MW, AbouChadi FE, Moustafa HE, Mekky N. Classification of ultrasound kidney images using PCA and neural networks. International Journal of Advanced Computer Science and Applications. 2015 Apr 1; 6(4):537.

[36] Subramanya MB, Kumar V, Mukherjee S, Saini M. "SVMbased CAC system for Bmode kidney ultrasound images", Journal of digital imaging. 2015 Aug 1; 28(4):44858.

[37] Rana S, Jain S, Virmani J.,"SVM Based characterization of focal kidney lesions from bmode ultrasound images", Research Journal Of Pharmaceutical Biological And Chemical Sciences. 2016 Jul 1;7(4):83746.

[38] Raja KB, Madheswaran M, Thyagarajah K.," Texture pattern analysis of kidney tissues for disorder identification and classification using dominant Gabor wavelet. Machine Vision and Applications", 2010 Apr 1; 21(3):287300.

[39] Priyanka, Dr. Dharmender Kumar, " Feature extraction and selection of kidney Ultra Sound Images using GLCM and PCA", International Conference on Computational Intelligence and Data Science, Procedia Computer Science 167 (2020) 1722–1731.

[40] Jyoti Verma, Madhwendra Nath, Priyanshu Tripathi, and K. K. Saini, "Analysis and Identification of Kidney Stone Using Kth Nearest Neighbour (KNN) and Support Vector Machine (SVM) Classification Techniques" , Pattern Recognition and Image Analysis, Vol. 27, No. 3, pp. 574–580 (2017).

[41] S. Selvaran & P. Rajendran, " Detection of Renal Calculi in Ultrasound Image Using Meta-Heuristic Support Vector Machine" Journal of Medical Systems 43: 300, (2019).

[42] Akkasaligar, P. T., and Biradar, S., " Diagnosis of Renal Calculus Disease in Medical Ultrasound Images" . IEEE International Conference on Computational Intelligence and Computing Research, 978-1-5090-0612-0/16 (2016).

[43] Elhoseny, M., Shankar, K. & Uthayakumar, J. " Intelligent Diagnostic Prediction and Classification System for Chronic Kidney Disease", Sci Rep 9, 9583 (2019). https://doi.org/10.1038/s41598-019-46074-2.

[44] I.A. Pasadana et al , " Chronic Kidney Disease Prediction by Using Different Decision Tree Techniques",2019 J. Phys.: Conf. Ser. 1255 012024.

[45] Chaudhuri, A.K., Sinha, D., Banerjee, D.K. et al. " A novel enhanced decision tree model for detecting chronic kidney disease" ,Netw Model Anal Health Inform Bioinforma 10, 29 (2021).

[46] Alaiad A, Najadat H, Mohsen B, Balhaf K ," Classification and association rule mining technique for predicting chronic kidney disease.", J Inf Knowl Manag , 2019(01):2040015.