Boosted Classifier and Features Selection for Enhancing Chronic Kidney Disease Diagnose

Made Satria Wibawa[1,a], I Made Dendi Maysanjaya[2,b], I Made Agus Wirahadi Putra[1,c]
[1)]Department of Information System, STIKOM Bali, Bali, Indonesia - 80226
[2)]Department of Informatics Education, Universitas Pendidikan Ganesha, Bali, Indonesia - 81116
[a)]satria.wibawa@stikom-bali.ac.id, [b)]dendi.ms@undiksha.ac.id, [c)]wirahadi@stikom-bali.ac.id

***Abstract*—Chronic kidney disease (CKD) is a disease caused by degeneration function of the kidneys. CKD is top ten leading causes of death in the world. There are two leading causes of CKD, i.e. diabetes and hypertension. When the symptoms become more severe, the disease can only be treated with dialysis and kidney transplantation. This disease can be treated if the diagnose is conducted appropriately and quickly. However, the signs and symptoms are often not specific. Because of that, diagnosis from medical personnel is may subjective and vary. This study developed machine learning method using ensemble learning and feature selection to improve the quality of CKD diagnosis. The CKD dataset was taken from UCI machine learning repository, it contain 400 instances. CKD dataset have 24 attributes including signs, symptoms and risk factors that might appear due to CKD. In this study, features were selected using a Correlation-based Feature Selection (CFS) and AdaBoost was used for ensemble learning to improve the detection of CKD. K-Nearest Neighbour algorithm (kNN), Naive Bayes and Support Vector Machine (SVM) was used as base classifier. Overall, the best result was achieved by combination of kNN classifier with CFS and AdaBoost, with 0.981 accuracy rate, 0.980 recall rate and 0.980 f-measure rate. Highest precision rate was achieved by the combination of Naive Bayes classifier with CFS and AdaBoost, with 0.981 precision rate.**

***Index Terms*—chronic kidney disease, CFS, kNN, Naive Bayes, SVM, AdaBoost, feature selection**

I. Introduction

Chronic Kidney Disease (CKD) is a term for various kinds of diseases that affect the structure and function of kidneys. CKD can be determined by the presence of kidneys damage called albuminuria or degradation of renal function such as glomerular filter rate (GFR) which less than 60 ml/min per 1.73 m^2 within a period of 3 months or more [1]

Kidney failure is the end-stage impact of the chronic kidney disease and the symptoms can be seen from the complications due to kidney function. When the symptoms become more severe, the disease can only be treated with dialysis and transplantation. The average number of chronic kidney disease incidents reached more than 200 cases per million people in a year in many countries in the world. Some countries such as the USA, Taiwan and some areas in Mexico have reached 400 cases per million people per year.

There are two leading causes of chronic kidney disease, i.e. diabetes and hypertension. In 2012, diabetes and hypertension are categorized among the world's top 10 causes of death. Diabetes was ranked ninth and killed 1.5 million people. Meanwhile, hypertension was ranked tenth and caused 1.1 million death [2].

CKD can be prevented by early detection and proper medical handling. However, because a relatively small number of nephrologist, not all patients with chronic kidney disease chronic get a proper diagnosis. The level of awareness of CKD in various health care providers still low even though they have experience more than 10 years in the medical field [3]. Therefore, an automated and accurate diagnosis method of CKD is necessary to assist medical personnel. Technology like computer-aided diagnosis (CAD) can help and assist health care providers in diagnosing chronic kidney disease. Various methods on CAD have been developed to help diagnosing disease symptom on the kidney, such as artificial neural network, fuzzy logic, neuro fuzzy, and so on.

A great deal of research has been conducted on the CAD for chronic kidney disease, but few experiments have been carried out to examine the combination of ensemble learning and feature selection in enhancing CKD classification. This study aims to enhance the quality of chronic kidney disease classification with feature selection method and ensemble learning. Those methods will be applied on the three classifiers, namely Naive Bayes, kNN, and SVM, that are used to classify the kidney into two classes (normal/without CKD and with CKD), by means WEKA application. In this work, the CKD dataset is taken from UCI Machine Learning Reposiroty [4].

The next section of the paper is organized as follows: Section II discusses research related to diagnosis of CKD, Section III discusses the research methodology, Section IV discusses the results and discussion and Section V discusses conclusions.

II. Literature Review

Chronic kidney disease is a term for degradation on kidney function in the certain period. The symptoms of this disease is obscure, but in the symptoms in general is malaise and reduced appetite. As previously mentioned, chronic kidney disease is determined by the presence of albuminuria. Degradation on renal function can be determined by the level of the glomerular filter rate (GFR), which is less than 60 mL/min per 1.73 m^2. Healthy adults have 125 mLmin per 1.73 m^2 of GFR level. Renal failure GFR of less than 15 mL/min per 1.73 m^2.

The level of GFR indicates stage of chronic kidney disease. Stage 1 is indicated by GFR level of over 90 mL/min per 1.73

m^2. Stage 2 is defined as a GFR of 60-89 mL/min per 1.73 m^2. Stage 3 is defined as a GFR between 30-59 mL/min per 1.73 m^2. Stage 4 is as a GFR between 15-29 mL/min per 1.73 m^2. Stage 5 is defined as a GFR of less than 15 mL/min per 1.73 m^2 [1]. Where the stage 1 is the most benign stage with minimal symptoms and stage 5 is the worst stages and can be led to death if not treated properly.

In addition to GFR, diagnosis of chronic kidney disease can be determined by examining at the symptoms and signs that might appears. Because the kidneys is highly adaptable organ, the signs and symptoms of chronic kidney disease may not be observable. These signs and symptoms can be appear such as causes, risk factors and complications. Some of them are hypertension, diabetes, older age, cardiovascular disease, anemia, the body fluids, urine, etc [5]. The signs and symptoms have correlation with chronic kidney disease. However the signs and symptoms of chronic kidney disease often are not specific, which means it can be caused by other diseases.

Signs and symptoms can be utilized for the chronic kidney disease detection . The use of machine learning methods have been successfully applied to biomedical fields, particularly to identify the disease and the risks appear based on those signs and symptoms. The main strength of the data mining is its interpretation ability on large datasets. The interpretation can be either the pattern or the knowledge of which is a generalization of the dataset. It is difficult for humans to determine the condition of the patients based on large and complex data. The subjectivity of human' decision can produce different conclusions. Therefore, data mining can be used to make a better and an objective decision regarding chronic kidney diagnose [6], [7].

Recent research on the diagnosis of chronic kidney disease using data mining has been done by some researchers. Some studies have used the UCI dataset [8]–[10] to diagnose chronic kidney disease. UCI chronic kidney disease dataset contain some missing values. This could be due to the fault of the receiver input, sensor error or reluctance on data resource. There are several ways to solve missing value problem, including instance removal [6], [9] or fill in missing value by statistical method such as median and mean or probability method [11].

As stated before, detection of chronic kidney disease can be conducted with data mining method, i.e. clustering [6] or classification [8]–[10]. Clustering in [6] is used to determine which patients need the most medical care based on Cluster Progression Score (CPS). Some famous classifier such as k-nearest neighbour, artificial neural network and naive bayes is used for the classification.

In addition to data mining, big data that is becoming the trend in data mining provides a great opportunity to be utilized in the medical field, particularly chronic kidney disease diagnosis. Big data involves the process of data mining from heterogeneous sources, autonomous with the complex relationships and growing. But the big data in diagnosis of chronic kidney disease has not been utilized [13].

UCI chronic kidney disease dataset have 25 attributes, including class attribute. Too many attributes can burden the classifier while making training data and increase computational time. There In addition to the problem of missing value, the number of attributes as much as 24 attributes (other than the class attribute) is considered relatively too much. Relatively many attributes can overload the klasifier in the manufacture of the model so that the computational time will also increase. There is also the possibility of not all attributes can contribute positively towards the calculation of the classification. One of the feature selection method that is often used is the wrapper. Although the wrapper takes a long computing but tend to produce a good classification results [12]. The wrapper has been used to reduce the number of attributes in the detection of chronic kidney disease in some research [8], [10]. The use of the wrapper on those research improved the results of classification of chronic kidney disease.

Beside feature selection, ensemble learning can be used to improve classification. Ensemble learning improve the confidence of classification by weighting various models generated by classifier and combining them to reach final model. Some of ensemble learning advantages are statistical reason, can handle well both of too little or too large volume data and it can divide the data into simpler space [16].

In this study, feature selection method and ensemble learning were proposed to enhance the quality of chronic kidney disease classification. Correlation-based Feature Selection (CFS) was used in feature selection and AdaBoost algorithm was used in ensemble learning. Three base classifiers were used to evaluating the effect of proposed method in enhancing classification result. Those classifiers were Naive Bayes, k-Nearest Neighbour and Support Vector Machine.

III. Methodology

This study generally was conducted into three main stage, namely feature selection, ensemble learning and classification. The original dataset with 25 attributes (including attribute classes) was reduced using Correlation-based Feature Selection (CFS). Three methods of classification, namely kNN, Support Vector and Naive Bayes were used as classifier. These classifier were used in ensemble learning for base classifier. Data and details of method are explained in the next section.

A. Data

This study use data from the UCI repository. The data collected from the Apollo Hospital, India by b. Jerlin Rubini. The number of instances in the dataset is 400, 250 instances was chronic kidney disease (ckd) and 150 instances was non-chronic kidney disease (notckd). There were 25 attributes, one of them was class attribute. The 24 attributes along with the abbreviation on the dataset were listed as follows: Age (age), Blood Pressure (bp), Specific Gravity (sg), Sugar (sg), Red Blood Cells (rbc), Cell (pc) Pussy, Puss Cell Clumps (pcc), Bacteria (ba), Blood Glucose Random (bgr), Blood Urea (bu), Serum Creatinine (sc), Sodium (sod), Potassium (pot), Hemoglobin (hemo), Packed Cell Volume (PVC), White Blood Cell Count (wc), Red Blood Cell Count (rc) , Hypertension

(htn), Diabetes Mellitus (dm), Coronary Artery Disease (cad), Appetite (appet), Pedal Edema (pe) and Anemia (ane) [4].

B. Feature Subset Selection

Feature subset selection aims to reduce computing time and improve the results of prediction of machine learning algorithms. This is done by reducing the features/attributes in a dataset that are considered unimportant or unable to contribute positively towards the classification, but it does not create new feature. Fewer features will reduces computing time [14].

Correlation-based Feature Selection (CFS) is suitable to be applied to multivariate data. CFS works by calculating the interaction between features. CFS evaluates a subset of features taking into account predictive capabilities of each level of redundancy among features and those features. The correlation coefficient is a feature used to calculate the correlation between the subset of features with feature classes and inter correlation among other features. CFS can be formulated on Equation 1.

$$r_{zc} = \frac{k\overline{r_{zi}}}{\sqrt{k + k(k-1)\overline{r_{zi}}}} \quad (1)$$

where r_{zc} is the correlation between product sum of the subset of features and feature classes; k is the number subset of features; r_{zi} is the average of the r_{zc} and r_{ii} is the average of the inter correlation among the subset fitur. [15]

C. AdaBoost

Ensemble learning combines some k models that have been produced from a classifier, $M_1, M_2, ..., M_k$ with goal to produce a combined classifier which better than M_*. If there are dataset of D, then training set will be produced as many as k, i.e. $D_1, D_2, ..., D_k$ in which D_i $(1 \leq i \leq k-1)$ is used for generated M_i classifier. Ensemble learning gives class prediction based on the voting of each model that are produced. Some example of ensemble learning are bagging, boosting, and random forest.

Boosting not only the select class prediction based on the number of voting the most, but also gives the weights of each M_i model. The weight is the value of an accuracy function on classifier. One of the popular boosting algorithms is AdaBoost, short for Adaptive Boosting. Suppose there are a D dataset with tuples that have classes a number of d, $(X_1, y_1), (X_2, y_2), ..., (X_d, y_d)$ in which y_i is a tuple of class X_i. AdaBoost will then assign weights for $1/d$ to each tuple training [11], [16]. AdaBoost algorithm can be seen on Algorithm 1.

Algorithm 1 AdaBoost

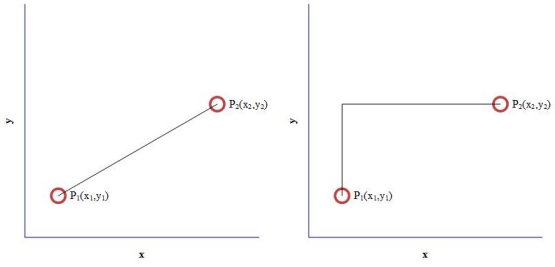
Input:
- D, a set of d class-labeled training tuples;
- k, the number of rounds
- a classification learning scheme

Output: A composite model

initialize the wight of each tuple in D to $1/d$;
for $i = 1$ to k **do**
  sample D with replacement according to the tuple wights to obtain D;
  user training set D_i to derive a model, M_i;
  compute $error(M_i)$, the error rate of M_i
  if $error(M_i) > 0.5$ **then**
    go back to step 3 and try again;
  end if
  for each tuple in D_i that was correctly classified **do**
    multiply the weight of the tuple by $error(M_i)(1 - error(M_i))$; //update weights
  end for
  normalize the weight of each tuple;
end for
return
To use ensemble to classify tuple, X:
initialize weight of each class to 0;
for $i = 1$ to k //for each classifier : **do**
  $w_i = log\frac{1-error(M_i)}{error(M_i)}$; //weight of the classifier's vote
  $c = M_i(X)$; //get class prediction for X from M_i
  add w_i to weight for class c
end for
return the class with largest weight;

D. kNN

A dataset consists of multiple instances, where each instance has more than one feature. This feature is a set of values that describe the characteristics of data. Data with the same class labels may have a similarity to each other. This similarity on cartesian coordinates is represented as the proximity of an instance. kNN exploit this principle, that the instance within a class will have affinity characteristics. Class of new instance can be determined by observing their nearest neighbor [17], [18].



(a) Euclidean distance (b) Manhattan distance

Fig. 1: Distance Equation

Closeness between instances can be measured by distance formula, for example Euclidean distance and Manhattan distance. Suppose there are two points x and y, Euclidean distance the shortest distance between two counting points and Manhattan distance to calculate the total distance that emerged from the second feature the points value. Illustration of the Euclidean and the Manhattan distance in two dimensions in a row can be seen in Figure 1a and Figure 1b.

On the Euclidean distance, the distance between two points x and y can be calculated with the Equation 2.

$$d(x,y) = \left(\sum_{i=1}^{m} |x_i - y_i|^2 \right)^{\frac{1}{2}} \tag{2}$$

On Manhattan distance, the distance between the point of the x and y can be calculated with the Equation 4.

$$d(x,y) = \sum_{i=1}^{m} |x_i - y_i| \tag{3}$$

In addition to both the Euclidean equations and Manhattan, there are also the Minkowsky formula shown in Equation 4 and Chebychev formula shown in Equation 5.

$$d(x,y) = \left(\sum_{i=1}^{m} |x_i - y_i|^r \right)^{\frac{1}{r}} \tag{4}$$

$$d(x,y) = \max_{i=1}^{m} |x_i - y_i| \tag{5}$$

This research used Euclidean distance. kNN have some advantages, such as it can generate a good accuracy, have relatively fast training time compared to other classifiers and overfitting case can be handled properly.

E. Naive Bayes

Naive Bayes is a simple classifier yet has a high scalability. This probabilistic model implements Bayesian theorem which assuming independence among variables/features. Because of its assumption on independency among variables, it is considered naive. But in real life, variables do not have the independency to one another. However, Naive Bayes is able to solve classification problem well .

Naive Bayes works as follows: suppose that D is a training set of a dataset and have the label class. Each instance is represented by a vector of attributes with n-dimensional, $X = (x_1, x_2, ...x_n,)$. Where there are a number of classes of m, which are $C_1, C_2, ..., C_m$. Assume if there is an instance of X, then classifier will predict a X to a class that has a highest posterior probability. So, the Naive Bayes will predict X instance into a class of C_i if and only if:

$$P(C_i \mid X) > \text{for} 1 \leq j \leq m, j \neq i \tag{6}$$

So, $P(C_i \mid X)$ will be maximized. Class C_i where $P(C_i \mid X)$ maximized is maximum posteriori hypothesis with Bayes' theorem:

$$P(C_i \mid X) = P(C_i \mid X) = \frac{P(X \mid C_i)P(C_i)}{P(X)} \tag{7}$$

Dataset which contain large attributes will have a high computational time to compute $P(X \mid C_i)$. Independency among attributes on Naive Bayes will play a role in this high computational time. Because of the independency among attributes, then:

$$\begin{aligned} P(X \mid C_i) &= \prod_{k=1}^{n} P(x_k \mid C_i) \\ &= (x_1 \mid C_i) \times P(x_2 \mid C_i) \times \cdots \times P(x_n \mid C_i) \end{aligned} \tag{8}$$

F. k-Folds Cross Validation

Data training and testing schema in this study used k-cross validation. On k-folds cross validation, each instances is used for training k-1 times and used as testing 1 time. The variance of classification result can be minimized by k-folds cross validation. The number of k used in the research was 10. K-folds cross validation schema can be ilustrated on Figure 2.

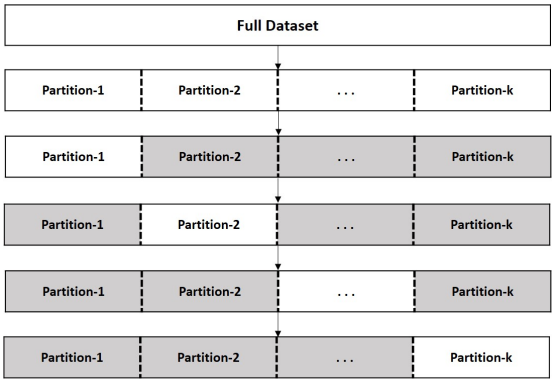


Fig. 2: k-Folds Cross Validation

G. Evaluation

To estimating classification performance, four parameters are used in this research. Those parameters are accuracy, precision, recall and f-measure. Accuracy describes the number of true prediction in and it is suitable for balanced data among class. However, because the data on CKD dataset is not balanced, precision, recall and f-measure will be used to assess classifier performance [11], [19].

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{9}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{10}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{11}$$

$$Fmeasure = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{12}$$

Precision describes how many true positive predicted data given all true predicted data. Recall describes how many true predicted data given all true real data and f-measure is harmonic mean from precision and recall. Accuracy, precision, recall and f-measure can be calculated consecutively in Equation 9, Equation 10, Equation 11 and Equation 3b.

IV. Result and Discussion

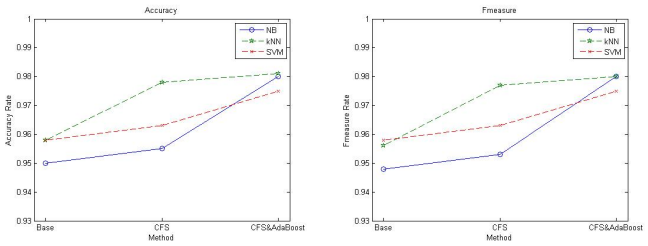
The use of the method of Correlation-based Feature Selection (CFS) removed 7 attributes. The result are 17 attributes i.e. Blood Pressure (bp), Sugar (sg), Albumin (al), Red Blood Cells (rbc), Blood Glucose Random (bgr), Blood Urea (bu), Serum Creatinin (sc), Sodium (sod), Potassium (pot), Hemoglobin (hemo), Packed Cell Volume (PVC), White Blood Cell Counts (wbcc), Hypertension (htn), Diabetes Mellitus (dm), Appetite (appet), Pedal Edema (pe), Anemia (ane).

TABLE I: Classification Result

Parameter	**Classifiers**		
	NB	**kNN**	**SVM**
	1st Method : Base		
Accuracy	0.950	0.958	0.958
Precision	0.941	0.949	0.958
Recall	0.960	0.966	0.958
F-measure	0.948	0.956	0.958
	2nd Method : CFS		
Accuracy	↑ 0.955	↑ 0.978	↑ 0.963
Precision	↑ 0.964	↑ 0.972	↑ 0.963
Recall	↑ 0.964	↑ 0.982	↑ 0.963
F-measure	↑ 0.953	↑ 0.977	↑ 0.963
	3rd Method : CFS + AdaBoost		
Accuracy	↑ ↑ 0.980	↑ ↑ 0.981	↑ ↑ 0.975
Precision	↑ ↑ 0.981	↑ ↑ 0.980	↑ ↑ 0.975
Recall	↑ ↑ 0.980	↑ ↓ 0.980	↑ ↑ 0.975
F-measure	↑ ↑ 0.980	↑ ↑ 0.980	↑ ↑ 0.975

The classification is conducted by three different methods, the first is classification by base classifier without a feature selection method and ensemble learning, the second is the result of the classification with feature selection but without the ensemble learning, the third is the result of the classification from selected features and ensemble learning. The arrow shows the changes of classification results from base classifier, ↑ sign means there is an increase in classification result and ↓ sign means there is a decrease in classification result. On the third method there are two arrows. The first or left arrow means changes from first method and the second/right arrow means a changes in classification from second method . The value of classification result can be seen in Table I and changes among classifier can be seen in Figure 3.

The accuracy on 1st method using base classifier are quite high, the lowest accuracy is 0.950. Accuracy rate in Naive Bayes is 0.950, both kNN and SVM accuracy rate is 0.958. 2nd method with CFS feature selection was able to increase the accuracy of base classifier. Accuracy rate in Naive Bayes, kNN and SVM are increased respectively by 0.005, 0.020 and 0.005. On the third method, features are selected by CFS and the classification is conducted with ensemble learning AdaBoost. Accuracy rate in three base classifier is increased on the third method. Accuracy rate in Naive Bayes is 0.980, kNN is 0.981 and SVM is 0.975. Highest accuracy rate is achieved on third method with kNN as a base classifier.



(a) Accuracy (b) F-measure

(c) Precision (d) Recall

Fig. 3: Evaluation Parameter Graph

Feature selection method and ensemble learning are successfully increase the precision rate on three base classifier. On base classifier, precision rate of Naive Bayes is 0.941, SVM is 0.949 and kNN is 0.958. Precision rate is increased on second method, Naive Bayes precision rate is 0.964, kNN is 0.972 and SVM is 0.963. The third method using CFS and AdaBoost also managed to increase the value of precision on the Naive Bayes became 0.981, kNN being 0.980 and SVM into 0.975.

In contrast with accuracy and precision, the recall rate is not increased in all method of classification. Classification on first method using Naive Bayes generates recall rate of 0.960, kNN is 0.966 and SVM is 0.958. Selected features with CFS is increased recall rate in Naive Bayes by 0.004, kNN by 0.016 and SVM by 0.005. In third method, precision rate from Naive Bayes, kNN and SVM consecutively are 0.964, 0.982 and 0.963. Precision rate in kNN is decreased from second method, but still higher than first method with base classifier.

The proposed method, namely AdaBoost and CFS was able to improve the classification results on all base classifier. The improvement of classification result can be seen in the four test parameters i.e. accuracy; precision; recall and f-measure. The best results of classification is obtained from third method with kNN and Naive Bayes as base classifier. The four parameter rate of the those two classifier is almost same, with the difference in the accuracy and precision. Accuracy rate in kNN is 0.981, Naive Bayes is 0.980 and precision rate in kNN is 0.980 and Naive Bayes is 0.981.

V. Conclusion

This study was designed to diagnose chronic kidney disease based on 24 attribute which includes symptoms, signs and risk factors of chronic kidney disease. The initial features

were selected using the method of Correlation-based Feature Selection (CFS). On the classification stage, AdaBoost was used for enhancing classification result. Three classifier, namely k-Nearest Neighbour (kNN), Support Vector Machine (SVM) and Naive Bayes were used to examining the effect of CFS and AdaBoost in enhancing classification result. There were three different classification method conducted. The first, classification conducted by only the base classifier. In the second method, classification was conducted after features was selected by CFS. In the third method, selected features were trained by AdaBoost learning. Classification was evaluated by using using four parameter, namely accuracy; precision; recall and f-measure. Based on four parameter of evaluation, CFS and AdaBoost were successful in improving chronic kidney disease diagnosis. There were increment in all classification methods. The best result was achieved by kNN classifier with 0.981 of accuracy rate.

REFERENCES

[1] A. S. Levey and J. Coresh, "Chronic kidney disease," *The Lancet*, vol. 379, no. 9811, pp. 165–180, 2012.

[2] W. H. Organization, "Fact Sheet : The Top Ten Causes of Death," 2012.

[3] L. C. Plantinga, D. S. Tuot, and N. R. Powe, "Awareness of Chronic Kidney Disease Among Patients and Providers," *Advances in Chronic Kidney Disease*, vol. 17, pp. 225–236, May 2010.

[4] M. Lichman, "UCI machine learning repository," 2013.

[5] "Symptoms and causes - Chronic kidney disease - Mayo Clinic."

[6] M. Lenart, N. Mascarenhas, R. Xiong, and A. Flower, "Identifying risk of progression for patients with Chronic Kidney Disease using clustering models," in *Systems and Information Engineering Design Symposium (SIEDS), 2016 IEEE*, pp. 221–226, IEEE, 2016.

[7] G. Dougherty, ed., *Medical Image Processing*. Biological and Medical Physics, Biomedical Engineering, New York, NY: Springer New York, 2011.

[8] A. Salekin and J. Stankovic, "Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes," pp. 262–270, IEEE, Oct. 2016.

[9] V. Kunwar, K. Chandel, A. S. Sabitha, and A. Bansal, "Chronic Kidney Disease analysis using data mining classification techniques," in *Cloud System and Big Data Engineering (Confluence), 2016 6th International Conference*, pp. 300–305, IEEE, 2016.

[10] N. Chetty, K. S. Vaisla, and S. D. Sudarsan, "Role of attributes selection in classification of Chronic Kidney Disease patients," in *Computing, Communication and Security (ICCCS), 2015 International Conference on*, pp. 1–6, IEEE, 2015.

[11] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques, Third Edition*. Waltham, Mass.: Morgan Kaufmann Publishers, 2012.

[12] M. Gtlein, E. Frank, M. Hall, and A. Karwath, "Large-scale attribute selection using wrappers," in *Computational Intelligence and Data Mining, 2009. CIDM'09. IEEE Symposium on*, pp. 332–339, IEEE, 2009.

[13] A. Batra, U. Batra, and V. Singh, "A Review to Predictive Methodology to Diagnose Chronic Kidney Disease," (India), pp. 2760–2763, IEEE, 2016.

[14] W. Awada, T. M. Khoshgoftaar, D. Dittman, R. Wald, and A. Napolitano, "A review of the stability of feature selection techniques for bioinformatics data," in *Information Reuse and Integration (IRI), 2012 IEEE 13th International Conference on*, pp. 356–363, IEEE, 2012.

[15] A. G. Karegowda, A. S. Manjunath, and M. A. Jayaram, "Comparative study of attribute selection using gain ratio and correlation based feature selection," *International Journal of Information Technology and Knowledge Management*, vol. 2, no. 2, pp. 271–277, 2010.

[16] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits and systems magazine*, vol. 6, no. 3, pp. 21–45, 2006.

[17] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," *Artificial Intelligence Review*, vol. 26, pp. 159–190, Nov. 2006.

[18] L. Jiang, Z. Cai, D. Wang, and S. Jiang, "Survey of Improving K-Nearest-Neighbor for Classification.," in *FSKD (1)*, pp. 679–683, 2007.

[19] I. H. Witten, E. Frank, and M. A. Hall, *Data mining: practical machine learning tools and techniques*. Burlington, MA: Morgan Kaufmann, 2011.