

PROGNOSIS OF CHRONIC KIDNEY DISEASE (CKD) USING HYBRID FILTER WRAPPER EMBEDDED FEATURE SELECTION METHOD

R.Parthiban¹, S.Usharani², D.Saravanan³,
D.Jayakumar⁴, Dr.U.Palani⁵, Dr.D.StalinDavid⁶, D.Raghuraman⁷

Associate Professor^{1,2,3,4}, Department of CSE, IFET College of Engineering, Villupuram,
Professor⁵, Department of ECE, IFET College of Engineering, Villupuram,
Assistant Professor⁶, Department of CSE, IFET College of Engineering, Villupuram,
Senior Assistant Professor⁷, Department of CSE, IFET College of Engineering, Villupuram,

Abstract:

Chronic Kidney disease (CKD) is a most predominant public health concern with increasing occurrence. CKD consists of an extensive variety of path physiological processes which will be experimental along with irregular function of kidneys and progressive decrease in Glomerular Filtration Rate (GFR). In CKD prediction various data mining methods play major important role and discovering the association among effective features in this stare canister lend a hand to detect or slow progression of this CKD disease. The information is serene from the patients' medical records. The major intention of this effort is introducing a Hybrid Filter Wrapper Embedded (HFWE) based Feature Selection (FS) to select optimal subset of features from CKD dataset. This HFWE-FS algorithm combines the procedure of filter, wrapper and embedded algorithm. Filter algorithm is executed based on the three major functions: Relief, One- R, Gain Ratio (GR) and Gini Index (GI). Wrapper algorithm is accomplished placed on the Improved Bat Algorithm (IBA) to choose analytical Attributes from the CKD dataset. Embedded algorithm is accomplished placed on the Support Vector Machine-t-statistics (SVM-t) to choose analytical attributes. The results of all feature selection algorithms are combined and named as HFWE- FS algorithm.

Keywords: *Chronic Kidney Disease (CKD), Improved Bat Algorithm (IBA), Feature selection (FS), Hybrid Filter Wrapper Embedded (HFWE), classification, Support Vector Machine (SVM).*

1 Introduction

Chronic Kidney Disease (CKD) progressively development and generally following months or years the kidney loses its functionality. In common it shouldn't be identified previous to it loses 25% of its functionality. The cause of renal failure shouldn't be identified by the patients because kidney failure shouldn't provide some symptoms originally. Kidney failure cures objective in the direction of the causes and slow down the proceed of the renal failure. If treatments are not sufficient, patient determination is in the final-stage of renal failure and the final cure is dialysis. Presently, four out of every 1000 person in the United Kingdom (UK) are suffering from CKD failure (Nordqvist 2016) and larger than 300,000 American patients in the final-stage of kidney disease continue to exist by dialysis Go et al (2004).

In accordance with the National Health Service, the kidney disease is found to be predominant in Africa, South Asia, when compared with other countries. This is primarily outstanding to the fact that until the disease progresses to the advanced level, detection of chronic complaint is not feasible. Hence nearby arise a necessitate for early detection of kidney failure, through which each kidney can be taken under control and consequently mitigate the risk of irreversible consequences. Prevention of renal failure and related complaints are possible (Nordqvist 2016). CKD can be detected through blood test which distinguishes measuring factors and thereby doctors can decide treatment process that reduces the progression rate Kathuria and Wedro (2016).

Due to the meagre availability of practicing nephrologists and their inability to entirely administer the patients with CKD, the burden of CKD management falls mainly on Primary Care Providers (PCPs). A recent study depicts that the awareness level of CKD by all categories of PCPs is unacceptably low and poor management on CKD knowledge among family practitioners. This is even common among those with one decade of clinical practice. Hence require for in finding an accurate, convenient and automated CKD detection method for clinical practice. But the challenge lies in reduction of dimensionality features.

Predictive class information are features independent. A comprehensive model for classification can be devised. The main focus of feature selection is achieving the best results of classification through recognition of the paramount detachment of appearance Plantinga et al (2010). Feature selection not only simplifies the data realization but also decreases the

over fitting problem and the size of data storage and thereby decreases the cost of train to attain greater accuracy Singh and Sivabalakrishnan (2015).

The taxonomy of feature selection methods consists of three groups that is in Figure 1. The filter method is independent of any learning algorithms which select Chao-Ton and Yang (2008). The performance improvement of classifier is attained by an appropriate feature selection and thereby reducing the computing time and by using optimized data in the dataset Kamari and Swanker (2011).

Owing to rapid performance and scalability, filter method remains admired among feature selection methods Villacampa, (2015),Karagandaet al (2010), Cho et al (2008). Wrapper method workout grades of feature sets that rely on the predicted power by using a classifier algorithm as a black box Ladha andDeepa(2011) . By executing analysis and instruction on that definite dataset, assessment of definite subset is reached and gains the space of all features of subsets Kumari and Swarnkar (2011). Decision support systems utilize dissimilar techniques to reduce the features aspect and classification algorithms to diagnose numerous kinds of infections.

Frequent accumulation of medical diagnosis results is extremely useful as its purpose is to mine the massive medical dataset to extract useful information. For the finding of CKD, Wrapper, filter and embedded based FS algorithms is used to diminish the length of features and therefore the commentary is prearranged as follows: Initially in Section 2 and the Section 3 details the proposed methodology. The Segment 4 observes the simulation test results for the used methods and its comparison. Finally, Section 5 culminates the commentary.

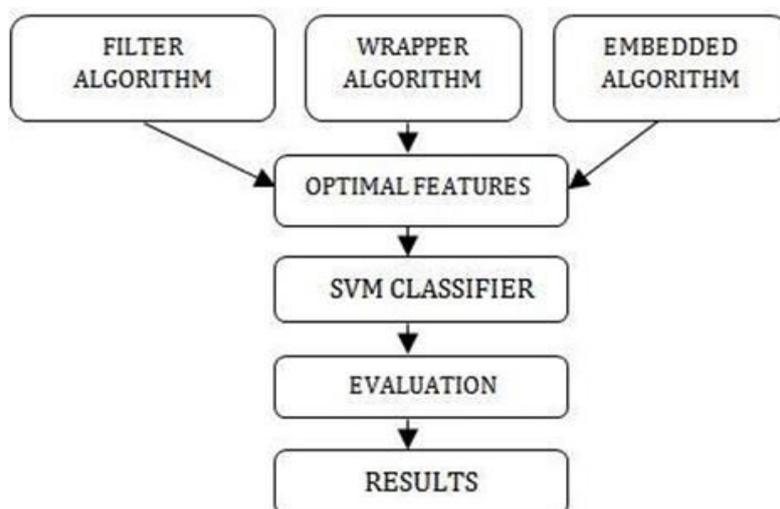


Fig 1. SVM classification with Hybrid Filter Wrapper Embedded (HFWE) algorithm

2 LITERATURE REVIEW

Misir et al (2017) explored whether CKD and non-CKD could be predicted with reasonable accuracy with fewer number of features. In this study, an intellectual scheme expansion loom that attempted one important feature selection technique to discover reduced features that explain the data set much better is used. Also to validate the reduced feature set two intelligent binary classification techniques is adopted. Performance evaluation is approved away by four important classification evaluation parameters. The consequences exposed to facilitate by concentrating more on the reduced features in the identification of CKD, and also considerable reduction in uncertainty and cost, by saving time.

Rubini and Eswaran (2015) introduced a novel CKD dataset with three classifiers such as Radial Basis Function (RBF) network, Multilayer Perceptron (MLP), and Logistic Regression (LR). The results were discussed in terms of prediction accuracy, type I error, type II error, type I error rate, type II error rate, sensitivity, specificity, F-score. The assess of accord between the arrangement made by the specialist and classifiers is represented as Kappa value. A comparative study of classifiers is done using performance accuracy.

Norouzi et al (2016) proposed a system for the prediction of renal failure timeframe and named as Adaptive Neuro Fuzzy Inference System (ANFIS) which used CKD depend on real clinical data. A decade of medical records of CKD diagnosed patients is taken for study. The entrance value of 15 cc/kg/min/1.73 m² of GFR was worn as the indicator of renal fault. A Takagi-Sugeno type ANFIS model was used to predict GFR values. For the purpose of predicting the model, they initially selected the variables such as age, sex, weight, underlying diseases, diastolic blood pressure, creatinine, calcium, phosphorus, uric acid, and GFR. To feed as input to the model the variables like Weight, diastolic blood pressure, diabetes mellitus as underlying disease, and current showed significant correlation with GFRs were selected. On comparing real data with predicted values, the study revealed that the ANFIS model is capable of accurately estimating GFR variations in all sequential periods and at extended prospect periods in spite of elevated qualms of human body and the active temperament of CKD progression.

O'Seaghdha et al (2012) developed a risk score to estimate an individual's absolute risk of incident CKD. The participants were observed for a decade to assess the development of CKD. To identify the risk factors of CKD a stepwise logistic regression was used. Later these were worn to create a risk score for the prediction of a decade CKD risk. Performance

measures evaluation were carried out by calibration and bigotry procedures. Further investigation was proposed for the utility of this score in identifying individuals in the community at high risk of chronic kidney disease.

Salekin and Stankovic (2016) introduced a machine learning wrapper method for the identification of a set of twelve attributes which exhibits CKD finding with high accuracy. The attributes were ranked according to their prognostic potential in detecting CKD and further reduce the prognostic attributes set to 10 using the LASSO regularization method. The experiment was conducted on a 400 individual's dataset, out of which 250 were detected for CKD. The results revealed that according to F1 measure recognition accurateness of 0.993 with 0.1084 root mean square error was attained. When compared with existing methods, a reduction of 56% mean square error was observed and the proposed method was capable of performing feature selection to resolve the most appropriate attributes for detecting CKD and ranking them according to their predictability.

Huang et al (2007) proposed a model by integrating Data Mining (DM) and Case-Based Reasoning (CBR) for the prognosis and diagnosis of a chronic disease. The model comprises of four main processes as follows: i) implicit meaningful rules are discovered by adopting DM techniques for examining health data, ii) prognosis of specific chronic diseases were done using extracted rules, iii) For diagnosis and treatment of chronic diseases CBR were employed and iv) For the expediency of chronic diseases information creating, organizing, refining, and sharing these processes were prolonged to work inside the classification. Di Noia et al (2013) describe a professional health examination center named MJ health screening center contributed the experimental data that was implemented and analyzed.

Polat et al (2017) study the Support Vector Machine (SVM) classifier to predict CKD. To predict the CKD, two major types of FS methods such as wrapper and filter algorithms were chosen to decrease the number of features of CKD dataset. Chen et al (2016) BFS engine in filter feature selection gives superior prophecy rate (98.5%) in the analysis of CKD when balanced to other feature selection algorithms.

3 PROPOSED METHODOLOGY

As Chronic Kidney Disease (CKD) development gradually, early recognition and successful dealing are the simply treatment toward decrease the death rate. Classification

algorithms are ahead The detection of classification methods are mainly depends on the selection of correct feature selection algorithms in the direction of decrease the dimension of datasets. In this work, Support Vector Machine (SVM) based machine learning algorithm was used in the direction of predicts CKD. To identify the CKD, three major types of FS algorithms namely, wrapper, filter and embedded methods were selected to decrease the number of features from CKD dataset. In filter approach, Relief, One-R, Gain Ratio (GR) and Gini Index (GI) were used. In wrapper approach, Improved Bat Algorithm (IBA) was used. Embedded algorithm is executed based on the Support Vector Machine- t-statistics (SVM-t) to choose analytical attributes from the CKD dataset.

Below Figure 2 shows the proposed schema of Hybrid Filter Wrapper Embedded (HFWE) for Predictive Attributes and SVM classification algorithm for the identification of the CKD.

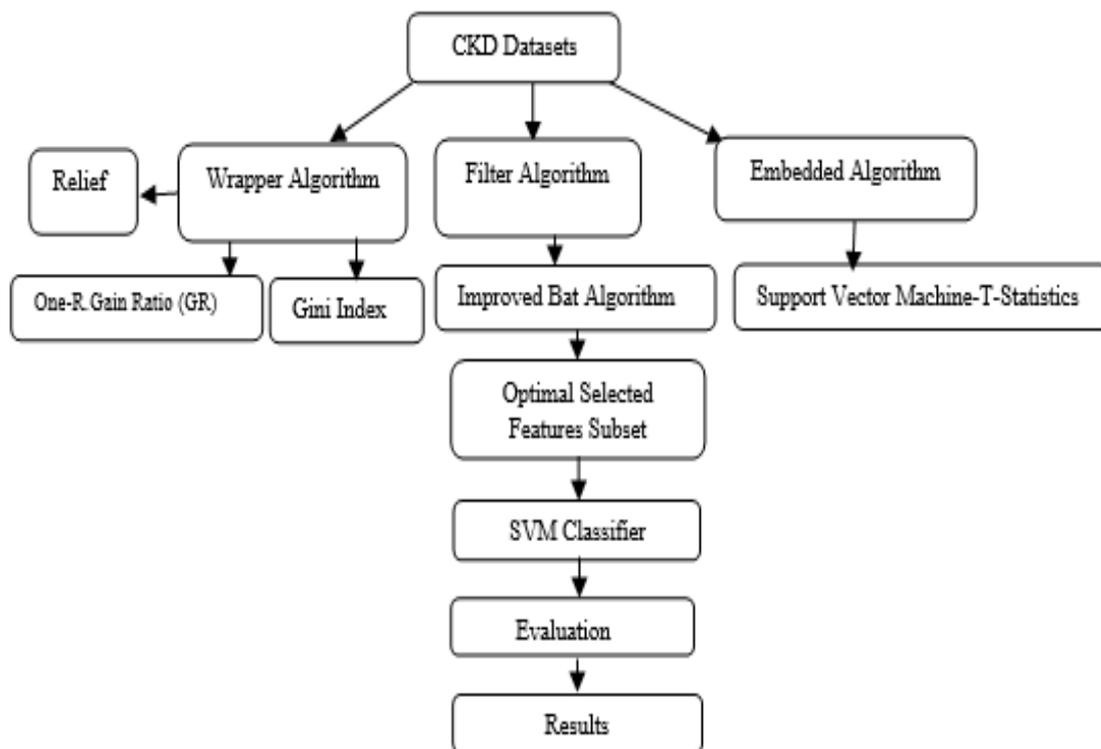


Fig 2. Proposed Hybrid Filter Wrapper Embedded (HFWE) Feature Selection for CKD prediction

3.1 Dataset Information

CKD is classified into Stages I–V as per the estimated Glomerular Filtration Rate (GFR) Levey and Coresh (2012). GFR is probable having arithmetic equations using serum

creatinine, age, sex, body size, ethnic origin, etc. (Am 2002). If the normal functionality of a kidney is besmirched to an extent, wastes can be erected up to elevated levels in the blood making sick.

3.2 Filter algorithm

The filter algorithms don't rely on any learning algorithm. The filter algorithm chooses the features whose ranks are the highest among them, and then the chosen subset features can be utilized for any predication algorithm.

3.3 Relief-F

Relief-F is an instance-dependent attribute selection algorithm which determines features by how fine its value differentiates examples with the purpose are from several groups however is comparable to each other. For each feature f_r , Relief-F chooses a random sample and k of its nearest neighbors from the similar class and each of several classes. Then f is scored as the sum of weighted variation in several classes and the same class. If f_r is differentially expressed, its determination show higher variation for instances from several classes, consequently it will obtain higher score (Holte1993).

$$SC_R(f_{r_i}) = \frac{1}{p} \sum_{t=1}^p \left\{ -\frac{1}{m_{f_{r_t}}} \sum_{f_{r_j} \in NH_{f_{r_t}}} d(f_{t,i} - f_{j,i}) + \sum_{y \neq y_{f_{r_t}}} \frac{1}{m_{f_{r_t}}} \frac{P(y)}{1 - P(y_{x_t})} \sum_{f_{r_j} \in NM(f_{r_t}, y)} d(f_{t,i} - f_{j,i}) \right\} \quad (1)$$

where $y_{f_{r_t}}$ is the class label of the sample f_{r_t} and $P(y)$ is the probability of an sample being from the class y . $NH(f_r)$ or $NM(f_r, y)$ represents a set of nearest points to sample f_r with the similar class of f_r , or a varied class (the class y), correspondingly. $m_{f_{r_t}}$ and $m_{f_{r_t}, y}$ are the sizes of the sets $NH(f_{r_t})$ and $NM(f_{r_t}, y)$, correspondingly. Generally, the size of both $NH(f_r)$ and $NM(f_r, y)$; $\forall y \neq y_{f_{r_t}}$, is set to a user defined constant k .

3.4 One-R

One-R is an easy algorithm Novakovic et al (2011), it creates one rule for each feature in the training data and then chooses the rule with the lesser error. It considers each and every one numerically valued attribute as continuous and makes use of a simple algorithm to separate the range of values addicted to a number of disjoint intervals. It handles missing data

by means of handling “missing” as a valid value. This is individual of the majority primitive methods. It creates easy rules depending on one characteristic simply. Though it is a smallest type of classifier, it is able to be helpful designed for determining a baseline results as a benchmark designed for further classification algorithms Lee et al(2011) .

3.5 Gain Ratio

The Gain Ratio is the non-symmetrical calculate with the purpose is developed in the direction of balance designed for the bias of the IG Komarasamy and Wahi (2012). GR is particular by means of equation (2)

$$GR = \frac{IG}{H(fr)} \tag{2}$$

In the equation (2) presents, when the feature Y has to be predicted, the Information Gain(IG) has to normalized by means of separating by means of the entropy of feature fr, and vice versa. Appropriate to this normalization, the Gain Ratio ideals constantly decrease in the series [0, 1]. An importance of Gain Ratio = 1 denotes with the principle of the information of feature entirely predict Y, and Gain Ratio (GR)= 0 means with the purpose there is no relation among Y and fr. The GR works well features with smaller amount values whereas the Information Gain (IG)

$$IG(fr, y) = H(fr) - H(fr|Y) \tag{3}$$

Entropy (H) is a calculate of the uncertainty related with an indiscriminate variable. H (fr) and H(fr/Y) is the entropy of fr and the entropy of successive observing Y, correspondingly.

$$H(fr) = - \sum_i P(fr_i) \log_2(P(fr_i)) \tag{4}$$

The highest value of IG is 1. A feature with a high IG is relevant. IG is determined independently for every feature and the features with the top-k values are chosen as the appropriate features. This filter-based FS algorithm doesn't remove redundant features.

$$H(fr|Y) = - \sum_j P(y_j) \sum_i P(fr_i|y_j) \log_2(P(fr_i|y_j)) \tag{5}$$

3.5 Gini Index(GI)

Gini index Kumar (2016)is supervised multivariate FS algorithm of the filter measure to determine for calculating a feature's capability to differentiate among classes. Specified C classes, GI of a feature are able to be determined as GIbe able to take the maximum value of 0.5 for a two-fold classification. The further appropriate features have smaller GI values. GI

of each feature is determined independently and the top k features with the smallest GI are chosen. Like IG, it moreover not removes unnecessary features.

$$GI(fr) = 1 - \sum_{i=1}^c [P(i|fr)]^2 \tag{6}$$

3.7 Wrapper algorithm

The wrapped search algorithm around the classifier picks up the space of all highlights of subsets.

3.8 Bat Algorithm (IBA)

Bat Algorithm, is executed based on the echolocation sounds of bats. Echolocation is characteristic sonar which bats make use to identify prey and to evade obstacles. These bats produce extremely noisy sound and pay attention for the echo with the purpose of jump back from the nearby features Yang (2010), Yang 2011). Consequently, a bat is able to calculate how far they are from a feature. Moreover, bats are able to differentiate the variety among an impediment and a quarry flat in total darkness Nakamura et al (2012). In order to selection of features from the dataset the bats algorithm follows some basic rules Komarasamy and Wahi(2012):

3.8.1 Initialization of Bat Population

Initial population of features is randomly created from original CKD dataset samples with data size d and number of bats n, by consideration of lower and upper boundaries.

$$x_{ij} = x_{min,j} + rand(0,1)(x_{max,j} - x_{min,j}) \tag{7}$$

where $i=1, 2, \dots, n$, $j=1, 2, \dots, d$, $x_{min,j}$ and $x_{max,j}$ are lower and upper boundaries for feature j correspondingly.

$f_i = f_{min} + (f_{max} - f_{min})\beta$	(8)
$v_i^t = v_i^{t-1} + (x_i^t - x_i^*)f_i$	(9)
$x_i^t = x_i^{t-1} + v_i^t$	(10)

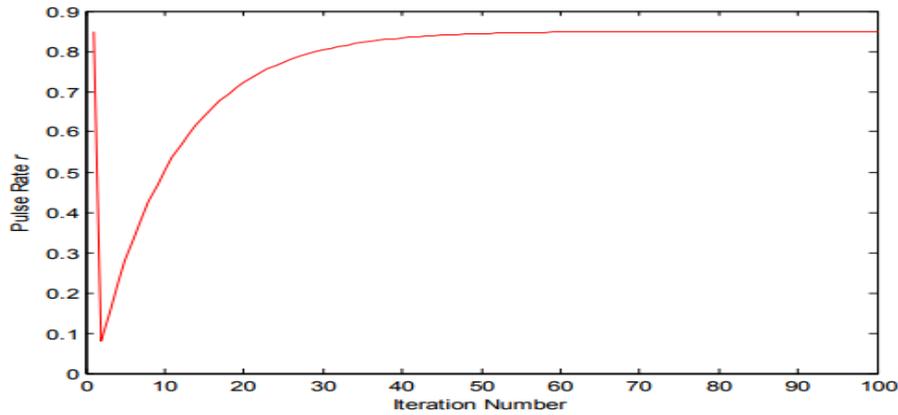


Fig 3. Pulse emission rate

$$x_{new} = x_{old} + \varepsilon \bar{A}^t \quad (11)$$

Where $t \in [0, 1]$ is a random number, \bar{A} is average loudness of each and every bats, ε denotes direction and intensity of random-walk.

$$w_{iter} = \frac{iter_{max} - iter}{iter_{max}}(w_{max} - w_{min}) + w_{min} \quad (12)$$

where $iter$ is current iteration value, $[iter]_{max}$ is maximum iteration number; w_{max} and w_{min} are greatest and least inertia weight factor respectively.

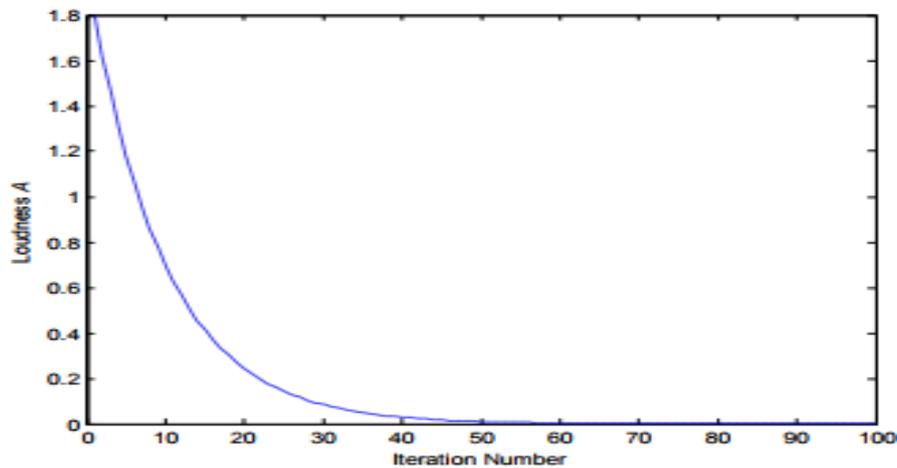


Fig 4. Loudness A

$$A_i^{t+1} = \alpha A_i^t \quad (13)$$

$$r_i^{t+1} = r_i^0 [1 - e^{-\gamma t}] \quad (14)$$

where α and γ are constants. r_i^0 and A_i^0 are factors which includes of random values and A_i^0 be able to characteristically be $[1, 2]$, while r_i^0 be able to characteristically be $[0, 1]$.

3.8.2 Improved Bat Algorithm (IBA)

BA is powerful algorithm at development however has some shortage at exploration Nakamura et al (2012), therefore it be able to straightforwardly get trapped in local minimum on the majority of the multimodal test functions. In order to solve this problem of standard BA, two modifications are applied to increase investigation and development ability of BA

3.8.3 Inertia Weigh Factor Modification

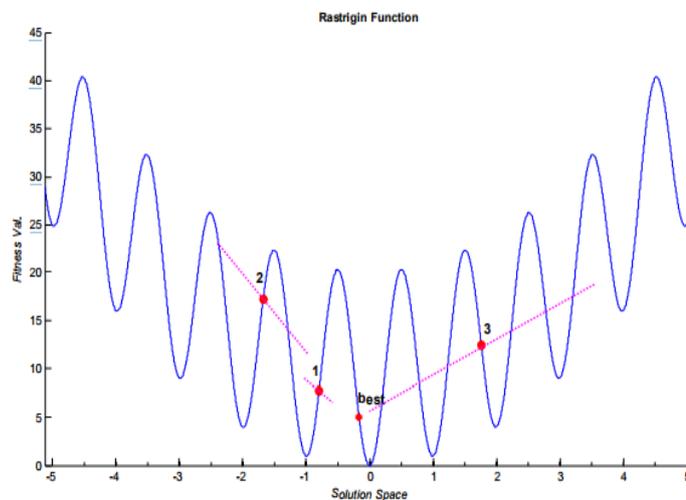


Fig 5. Distribution of frequency

Linear decreasing inertia weight factor the result of earlier velocity progressively decreases. Consequently, development rate of BA progressively enhances as iterations proceed.

3.8.4 Adaptive Frequency Modification

Any randomly created frequency value is allocated a feature selection solution in BA and this frequency value will have same effect to each and every one dimensions of such feature selection solution. The variation between all the features of a solution has no sense at this point. This formation decreases local search results of BA. However, in IBA each feature of a solution is allocated a frequency from f_{min} to f_{max} individually.

$$w_{iter} = \frac{iter_{max} - iter}{iter_{max}}(w_{max} - w_{min}) + w_{min} \quad (15)$$

where $iter$ is current iteration value, $[iter]_{max}$ is maximum iteration number; w_{max} and w_{min} are greatest and least inertia weight factor respectively.

$$\text{diff}_j = \sqrt{((x_{ij} - x_{j^*})^2)} \quad (16)$$

$$\text{Range} = \max(\text{diff}) - \min(\text{diff}) \quad (17)$$

$$f_j = f_{\min} \left[\left[\frac{(\max(\text{diff}) - \text{diff}(j))}{\text{range}} \right] * (f_{\max} - f_{\min}) \right] \quad (18)$$

$$v_{ij}^t = v_{ij}^{(t-1)} + (x_{ij}^t - x_{j^*}^t) f_j \quad (19)$$

3.9 Embedded algorithm

Embedded algorithm is executed based on the Support Vector Machine- t-statistics (SVM-t) to choose analytical attributes from the CKD dataset. SVM make use of only the information of support vectors in the direction of create the maximal partition hyper plane and find the classes for every CKD dataset. The support vectors, the position of closest CKD dataset points among two classes, participate significant role for feature selection. The proposed SVM-t method makes use of the most significant feature subset (i.e., SVs) of the CKD dataset points to create the feature selection criteria (19). In other words, we use the algorithm of SVM as a sampling algorithm for data points of two classes concurrently. The standard two-sample t-statistic is used as a replacement statistic in the direction of evaluates the important variation among two classes. So, with the difference of samples, we are able to classify the most important dissimilarity for precise genes between the closest data points:

$$|t_j| = \left| \frac{(u_{j^+} - u_{j^-})}{\sqrt{\left(\frac{(s_{j^+})^2}{n^+} + \frac{(s_{j^-})^2}{n^-} \right)}} \right| \quad (20)$$

EXPERIMENTAL RESULTS

Dataset Information

CKD is open into Stages I–V as per the estimated Glomerular Filtration Rate (GFR) shown in Table 1 Levey and Coresh (2012). If the usual performance of a kidney is tainted to an extent, wastes can build up to high levels in the blood making suffer (Am 2002). The confusion matrix is used to narrate the performance of classification algorithms by evaluating the performance metrics.

True Positive (TP) - Indicates optimistic instances properly categorized as affirmative outputs, True Negative (TN) - Indicates pessimistic instances properly categorized as pessimistic outputs, False Positive (FP) - Indicates pessimistic instances wrongly categorized as optimistic outputs and False Negative (FN) - Indicates optimistic instances wrongly

categorized as pessimistic output. Classification Accuracy - Indicates the ability of classifier algorithm to diagnose of classes of dataset

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{TN} + \text{FN}) \times 100 \quad (21)$$

$$\text{Recall} = \text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN}) \times 100 \quad (22)$$

Specificity relates to the test's ability to correctly detect patients without a condition.

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP}) \times 100 \quad (23)$$

Precision also called positive predictive value is the fraction of relevant instances among the retrieved instance

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \times 100 \quad (24)$$

F-measure is the harmonic mean of precision and recall is measured as follows:

$$\text{F-measure} = 2 \times (\text{Recall} \times \text{Precision}) / (\text{Recall} + \text{Precision}) \times 100 \quad (25)$$

Specificity relates to the test's ability to correctly detect patients without a condition.

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP}) \times 100 \quad (26)$$

Precision also called positive predictive value is the fraction of relevant instances among the retrieved instance

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \times 100 \quad (27)$$

Table 1. Performance metrics vs. classifiers

Methods	Results(%)					
	Sensitivity	Specificity	Precision	F-measure	Accuracy	Error rate
NB- HFWE- FS	90.48	72.73	88.79	89.62	85.23	14.77
ANN- HFWE-FS	91.59	74.42	89.91	90.74	86.67	13.33
SVM - HFWE-FS	95.45	87.50	95.45	95.45	93.33	6.67
NB	75.00	33.33	81.82	78.26	66.67	33.33
ANN	76.19	55.56	80.00	78.05	70.00	30.00
SVM	77.27	62.50	85.00	80.95	73.33	26.67

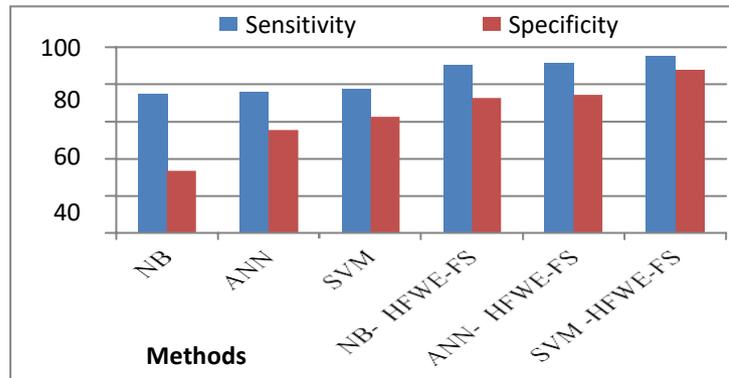


Fig 6. Classifiers vs. metrics (sensitivity and specificity)

Fig 6 shows the results are measured in terms of the feature selected from the HFWE-FS algorithm; it concludes that these three classifiers perform better under HFWE-FS algorithm.

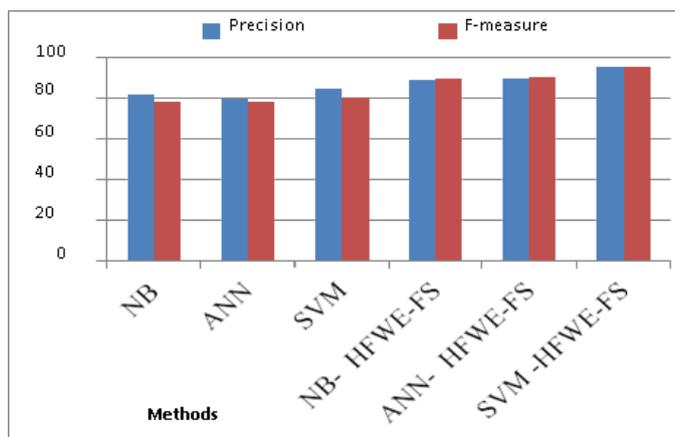
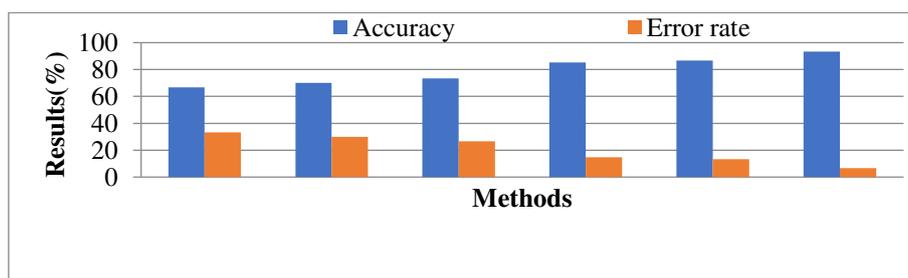


Fig 7. Classifiers vs. metrics (Precision and F-measure)

Fig 7 shows the results of the precision and recall metrics with respect to three different classifiers such as the NB, ANN and SVM. HFWE-FS and NB- HFWE-FS classifiers respectively.



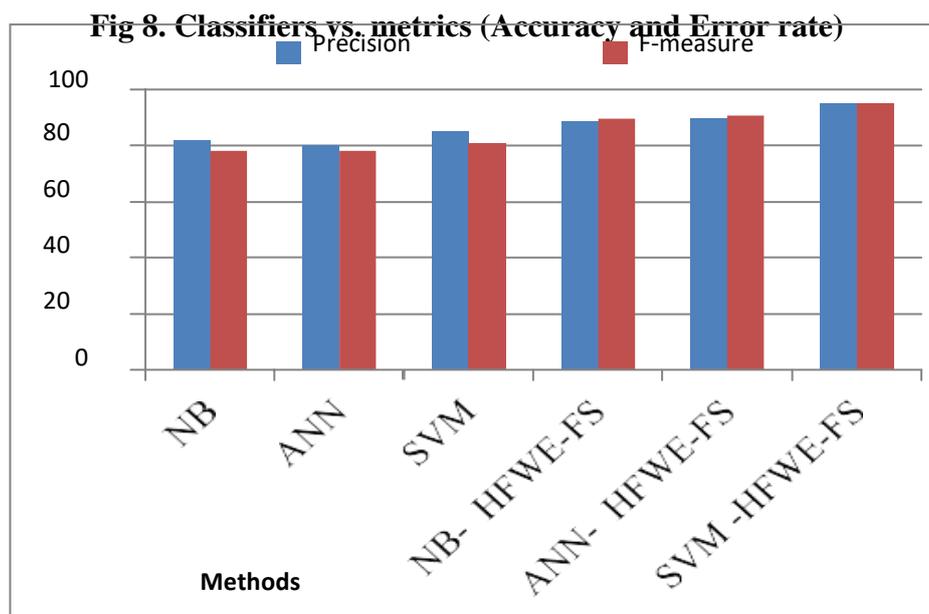


Fig 8 shows the results of the accuracy and error rate metrics with respect to three different classifiers such as the NB, ANN and SVM. The results are measured using the features selected from HFWE-FS algorithm; it proposed HFWE-FS algorithm with three classifiers produces higher results than traditional classifiers.

4 CONCLUSION AND FUTURE WORK

This paper has introduced a novel Hybrid Filter Wrapper Embedded (HFWE) based Feature Selection (FS) algorithm to select optimal subset of features from the datasets to predict CKD datasets. For the analysis of CKD, Wrapper, filter and embedded based FS algorithms is worn to decrease the breadth of features and consequently SVM has been used for categorization of features. Filter algorithm is executed based on the three major functions: Relief, One-R, Gain Ratio (GR) and Gini Index (GI). Wrapper algorithm is executed based on the Improved Bat Algorithm (IBA) to choose analytical Attributes from the CKD dataset. Embedded algorithm is carried out depend on the Support Vector Machine-t-statistics (SVM-t) to choose analytical attributes from the CKD dataset. The popularly known SVM classifier have been followed for the validity of the reduced feature set. Executed an evaluation on a dataset of 400 patients, 250 among them have early stage of CKD. CKD samples were collected from University of California Irvine (UCI) machine learning repository. According to the UCI data set, there are 24 attributes for predicting CKD or non-CKD. The target of this effort is to discover whether it is possible to predict CKD or non-CKD with reasonable accurateness by selected features from HWFFS algorithm. Performances were determined in

terms of six important classification evaluation parameters. However, the proposed SVM-HFWE-FS algorithm produces higher accuracy results of 93.33% which is 6.66% and 8.1% higher when compared to ANN- HFWE-FS and SVM- HFWE-FS algorithm. This dataset contains some noisy and missing values. Therefore, a classification algorithm is required with the capability of dealing with missing and noisy values are kept as future work.

ACKNOWLEDGMENT

The authors are thankful to everyone, who has supported them along the way. They also grateful to their family members and friends who have provided them through moral and emotional support in their life.

Compliance with Ethical Standards

Conflict of Interest

The authors declared that they have no conflict of interest.

Funding

There is no funding source for this work.

References

1. Am J National Kidney Foundation. K/DOQI clinical practice guidelines for chronic kidney disease: Evaluation, classification, and stratification. *Kidney Dis*, 2002; 392 Suppl1:S1-266.
2. Anderson, J; Glynn LG. Definition of chronic kidney disease and measurement of kidney function in original research papers: A review of the literature. *Nephrol Dial Transplant*, 2011 ; 26:2793-2798.
3. Chao-Ton, S., Yang, C.-H. Feature selection for the SVM: an application to hypertension diagnosis, *Expert Syst. Appl.* 2008; 34:754– 763.
4. Chen, Z., Zhang, X., and Zhang, Z. Clinical risk assessment of patients with chronic kidney disease by using clinical data and multivariate models”. *Int. Urol. Nephrol.* 2016; 48:2069–2075.
5. Cho, B.H., Yu, H., Kim, K.-W., Kim, T.H., Kim, I.Y., and Kim, S.I.(2008). Application of irregular and unbalanced.

6. D.Stalin David D, "Textures and intensity histogram based retinal image classification using hybrid color structure descriptor", Biomedical and Pharmacology Journal, 2018; Vol.11 issue 1, pp.577-582.
7. Go, A.S., Chertow, G.M., Fan, D., McCulloch, C.E., Hsu, C.-y. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization". 2004; N. Engl. J. Med.
8. D.Stalin David D, "Parasagittal Meningioma Brain Tumor Classification System based on MRI Images and Multi Phase level set Formulation", Biomedical and Pharmacology Journal, 2019; Vol.12, issue 2, pp.939-946.
9. Huang, M.-J., Chen, M.-Y., Lee, S.-C. Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis. Expert Syst. Appl. 2007; 32:856–867.
10. Karegowda, A.G., Jayaram, M.A.,Manjunath, A.S..Feature subset selection problem using wrapper approach in supervised learning. Int. J. Comput. Appl. 2010; 1(7):13–17.
11. Kathuria, P., and Wedro, B .Chronic kidney disease quick overview. IOP Publishing emedicine health, http://www.Emedicinehealth.com/chronic_kidney_disease/page2_em.htm#chronic_kidney_disease_quick_overview.. 2016.
12. Komarasamy, G; Wahi, A. "An optimized K-means clustering technique using bat algorithm". European Journal of Scientific Research, 2012; 84(2), pp.263-273.
13. Kumar, M.,(2016).Prediction of Chronic Kidney Disease Using Random Forest Machine Learning Algorithm". International Journal of Computer Science and Mobile Computing, 5(2), pp.24-33.
14. Kumari, B., Swarnkar, T. Filter versus wrapper feature subset selection in large dimensionality micro array: a review, International Journal of Computer Science and Information Technologies. 2011; 2(3):1048–1053.
15. Ladha, L., Deepa, T . Feature selection methods and algorithms. International J. Comput. Sci. Eng. 2011; 3(5):1787–1797, 2011.
16. D. Stalin David, "A comprehensive survey of security mechanisms in healthcare applications," International Conference on Communication and Electronics Systems (ICCES), Coimbatore, 2016, pp. 1-6.
17. Levey AS.; Coresh J. Chronic kidney disease. Lancet ; 2012; 379:165-80.

18. Misir, R., Mitra, M. ;Samanta, R.K.(2017). A reduced set of features for chronic kidney disease prediction. *Journal of pathology informatics*.
19. Nakamura,R ; Pereira,L; Costa,K; Rodrigues,D; Papa,J; Yang,X.S. BBA: A Binary Bat Algorithm for Feature Selection,” in Proc. 25th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), 2012; Aug. 22-25, pp. 291-297.
20. Nickabadi,A; Ebadzadeh,M; Safabakhsh,R .A novel particle swarm optimization algorithm with adaptive inertia weight,” *Applied Soft Computing*, 2011; vol. 11, pp. 3658-3670.
21. D. Stalin David, A new expert system based on hybrid colour and structure descriptor and machine learning algorithms for early glaucoma diagnosis. *Multimed Tools Appl*, 2018; 79, Issue 7-8, 5213–5224. <https://doi.org/10.1007/s11042-018-6265-1>.
22. Norouzi, J., Yadollahpour, A., Mirbagheri, S.A., Mazdeh, M.M.; Hosseini, S.A. Predicting renal failure progression in chronic kidney disease using integrated intelligent fuzzy expert system. 2016; *Computational and mathematical methods in medicine*.
23. Novakovic,J; Strbac, D.; Bulatovic. Toward optimal feature selection using ranking methods and classification algorithms,” *Yugoslav Journal of Operations Research*, 2016; vol. 21, no. 1, pp. 119-135.
24. O'Seaghda, C.M., Lyass, A., Massaro, J.M., Meigs, J.B., Coresh, J., D'Agostino, R.B., Astor, B.C. and Fox, C.S. A risk score for chronic kidney disease in the general population. *The American journal of medicine*, 2012; 125(3), pp.270-277.
25. Plantinga, L.C; D. S. Tuot, D.S; Powe, N. R “Awareness of chronic kidney disease among patients and providers, *Advances in chronic kidney disease*, 2010; vol. 17, no. 3, pp. 225–236.
26. Polat, H., Mehr, H.D.;Cetin, A. Diagnosis of Chronic Kidney Disease Based on Support Vector Machine by Feature Selection Methods. *Journal of medical systems*, 2017; 41(4), pp.1-11.
27. Rubini, L.J;Eswaran, P.(2015). Generating comparative analysis of early stage prediction of Chronic Kidney Disease. *International OPEN ACCESS Journal Of Modern Engineering Research*, 5(7), pp.49-55.
28. Salekin, A.;d Stankovic, J. Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes. *IEEE International Conference on Healthcare Informatics (ICHI)*, 2016; pp. 262-270.

29. Singh, R.K., Sivabalakrishnan, M. Feature selection of gene expression data for cancer classification: a review, *Procedia Computer Science*. 2015; 50:52–57.
30. D. Stalin David, Robust Classification of BrainTumor in MRI Images using Salient Structure Descriptor and RBF Kernel-SVM, *TAGA Journal of Graphic Technology*, 2018; Volume 14, Issue 64, pp.718-737.
31. Stalin David D, Saravanan D, 2020, ‘Multi-perspective DOS Attack Detection Framework for Reliable Data Transmission in Wireless Sensor Networks based on Trust’, *International Journal of Future Generation Communication and Networking* , Volume 13, Issue 4, PP.1522–1539.
32. J. K. S and D. S. David, "A Novel Based 3D Facial Expression Detection Using Recurrent Neural Network," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2020, pp. 1-6, doi: 10.1109/ICSCAN49426.2020.9262287.
33. R.Parthiban, Dr.K.Santhosh Kumar, Dr.R.Sathya, D.Saravanan,"A Secure Data Transmission And Effective Heart Disease Monitoring Scheme Using Mecc And Dlmnn In The Cloud With The Help Of Iot", *International Journal of Grid and Distributed Computing*, ISSN: 2005 – 4262, Vol. 13, No. 2, (2020), pp. 834 – 856.
34. R.Bhavya, G.I.Archanaa, D.Karthika, D.Saravanan," Reflex Recognition of Tb Via Shade Duplicate Separation Built on Geometric Routine", *International Journal of Pure and Applied Mathematics* 119 (14), 831-836.
35. D Saravanan, R Bhavya, GI Archanaa, D Karthika, R Subban," Research on Detection of Mycobacterium Tuberculosis from Microscopic Sputum Smear Images Using Image Segmentation", 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICIC).
36. D Saravanan, R Parthiban," Automatic Detection of Tuberculosis Using Color Image Segmentation and Statistical Methods", *International Journal of Advance Research in Science and Engineering*, Volume 6, Issue 10.
37. U.Palani, D.Saravanan, R.Parthiban, S.Usharani," Lossy Node Elimination Based on Link Stability Algorithm in Wireless Sensor Network", *International Journal of Recent Technology and Engineering (IJRTE)*, Volume 7, Issue 6S5.
38. S.G.Sandhya, D.Saravanan, U.Palani, S.Usharani," Handover Priority to the Data at Knob Level in Vanet", *International Journal of Recent Technology and Engineering (IJRTE)*, Volume 7, Issue 6S5.

39. D.Saravanan R. Parthiban, U.Palani S.G. Sandhya, " Sheltered and Efficient Statistics Discrimination for Cluster Based Wireless Antenna Networks", International Journal of Recent Technology and Engineering (IJRTE), Volume 7, Issue 6 S5.
40. D.Saravanan¹, Dr.K.Santhosh Kumar², R.Sathya³, U.Palani⁴, "An Iot Based Air Quality Monitoring And Air Pollutant Level Prediction System Using Machine Learning Approach – Dlmnn", International Journal of Future Generation Communication and Networking, Vol. 13, No. 4, (2020), pp. 925–945.
41. Raghu Raman D, Saravanan D, Nivedha R, "An Efficacious E-Portal for Rancher to Buy Seeds and Humus", International Journal of Recent Technology and Engineering (IJRTE), Volume-8, Issue-1 S5, June 2019.
42. M.Sudha, D.Saravanan, S.Usharani, " Security Improvement of Dropper Elimination Scheme for IoT Based Wireless Networks", International Journal of Engineering Trends and Technology (IJETT), Volume-45 Number3 -March 2017.
43. K. Gayathri, D.Saravanan, " An Innovative IOT security Enhancing Schema Based Gamed Theory Decryption Percentage", International Journal of Advanced Research in Science and Technology, Volume 6, Issue2, 2017, pp. 666-671.
44. E.Kowsalya, D.Saravanan, R.Parthiban, " Energy Aware Resource Allocation for Throughput Maximized IOT Network", International Journal of Computer Trends and Technology (IJCTT) – Volume 45 Issue 2- March 2017.
45. Yang, X.S. A New Metaheuristic Bat-Inspired Algorithm in Nature Inspired Cooperative Strategies for Optimization, J. R. Gonzalez et al., Eds., Springer Press, 2010; vol. 284, pp. 65-74.
46. Yang, X.S. Bat algorithm for multi-objective optimization, International Journal of Bio-Inspired Computation, 2011; vol. 3, no.5, pp.267-274.
47. Stalin David D, Saravanan D, "Enhanced Glaucoma Detection Using Ensemble based CNN and Spatially Based Ellipse Fitting Curve Model", Solid State Technology, Volume: 63 Issue: 6.
48. 32. Stalin David D, Saravanan D, Dr.A.Jayachandran " Deep Convolutional Neural Network based Early Diagnosis of multi class brain tumour classification system", Solid State Technology, Volume: 63 Issue: 6.