# ASSESSING THE EFFECT OF THE LIBSVM CLASSIFIER ON LINEAR AND POLYNOMIAL FUNCTION AS A RESULT OF THE RESEARCH PROGNOSIS FOR THE HARNESSING OF STATISTICS ON CHRONIC KIDNEY DISEASE – A CASE STUDY OF CKD PATIENTS

**Rajesh S. Walse**, School of Computational Sciences (RC), Swami Ramanand Teerth Marathwada University, Nanded, CDT, Pusad; M.A.F.S.U, Nagpur, India
**Dr. Gajanan D. Kurundkar** Department of Computer Science,Shri Gurubuddhi Swami Mahavidyalaya, Purna Dist. Parbhani, S.R.T.M.University, Nanded, India,
**Dr. Sachin M. Narangale** School of Media Studies, S.R.T.M.University, Nanded, India
rajeshwalse@gmail.com, gajanan.kurundkar@gmail.com, sachin.narangale@gmail.com

**Abstract**: For this experiment, CKD patient data was analyzed using supervised LibSVM classifier algorithm with two functions of Linear and Polynomial. Standards have been used to define which function gets the best statistics for CKD patients. Using WEKA tool, we used two methods for linear and polynomial function, initially select cluster mode, using training data set of 66 percent and 70 percent with application of Cross Validation Folds=10 and number of attributes using 1,3,5,10,15 and 24 to find the best module for precision. Many novel features are observed, the best and fastest classifier module of LibSVM has been found and the best accuracy for the linear function of Correctly Classified Instances has been found to be 94 percent and 95 percent and the curve value of the receiver operating characteristics has been increased and found to be close to 1 percent, which was then detected with the highest accuracy. This technique has been shown to be an efficient way to evaluate CKD using this model and to forecast. The research paper's main objective is to evaluate analysis of the Linear function and polynomial function classifier LibSVM. Which prove a mechanism for Linear function classifier LibSVM provides the best precision compared to Polynomial function. Increased accuracy of Linear function is suggested, based on increased attributes, and found best model.

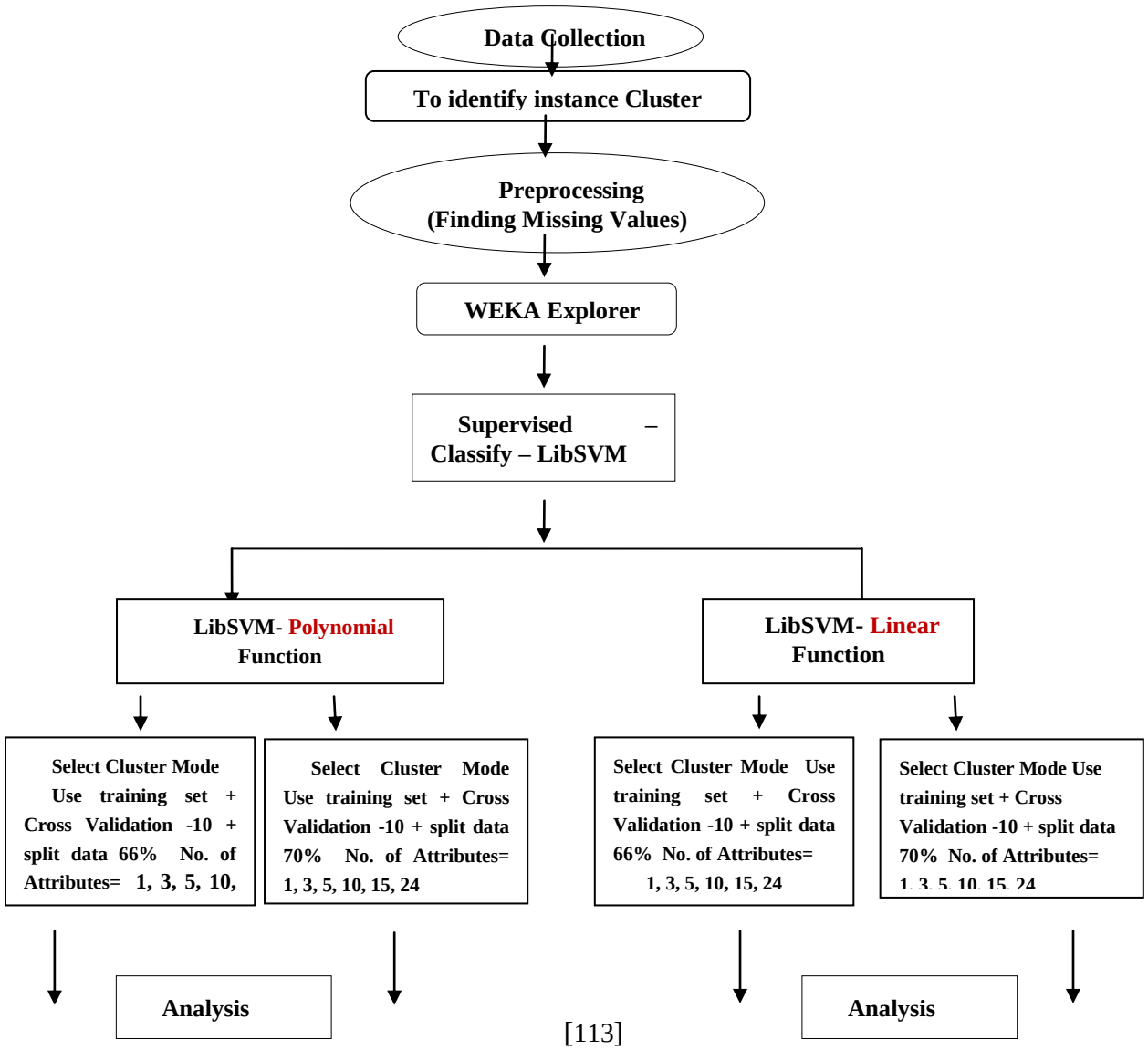
**Keywords:** Supervised, Classification, LibSVM, WEKA, Linear, Polynomial, CKD

## I. INTRODUCTION

The researcher tries to fill the gap of earlier research and this type of study is filled the gap by using this type of methodology adopted and enhancement and correct outcome. The techniques of data mining are the process of identifying the hidden patterns of large and tedious data. It can play a vital role in decision-making on broad numbers, and not just Not only the problems related to agriculture but also health. Bharara et al.[3] reviewed machine learning techniques to extract for business practices. Ariff et al.[2] researched the hierarchical method of livestock management based on RFID. Jinyin[7] is the fast determination clustering algorithm performed by a novel cluster core. DilliArasu and Thirumalaiselvi[1] dealt with novel techniques for imputation Efficient kind of kidney disease prediction for patients. ZouChuan et al.[4] performed an integrated study of traditional Chinese care at Guangdong provincial hospital. Guangzhou and discuss clustering study for peritoneal dialysis patients with syndrome evolution Kunwar et al.[9] researched and examined Chronic for classification strategies in terms of persistent Kidney Disease harnessing of data mining. Sabri[6] used data mining techniques to classify knowledge about the customers. WEKA used by Kumar and Lhatri[10] is used for classification of medical related data and for seeking prediction of early disease. Khanna[10], NCBI[12] conducted a review into Dialysis patients of economics. J Nephrol[13] examined the advent of chronic kidney disease in India, and where are we heading? Ubolthamet al.[11] did an acute kidney injury diagnostic analysis using the KDIGO guideline method. This paper Study conducted on chronic kidney disease patients based on their relationship characteristics, nowadays chronic kidney disease patients in India are rising daily. Its eating habits

and other health issues. Nonetheless, from the last ten years , the number of CKD patients has increased tremendously in the Indian Journal of Nephrology et al.[12], thus in the future this form of research that will enable doctors or the medical industry to predict CKD rather than CKD patients on the basis of their previous data available to doctors on the basis of their other health parameters. To decrease the growth rate of patients with CKD and to monitor more kidney damage. We used secondary data to test the results, and it is retrieved from the UCI machine learning repository[14]. With a increasing lifespan and a prevalence of lifestyle disease, Jnephrol[13] has seen a substantial 30 percent rise in the widespread incidence of CKD in the last decade. The linear method of the LibSVM classifier algorithm is the best and provides the most reliable and correct results for 66 percent and 70 percent of the split value (training on a portion of the data and checking on the rest) accuracy is 94 percent accuracy for 66 percent split trained data and 95 percent accuracy for 70 percent split trained data and 1 percent ROC value. The analysis limitation is comparative evaluation of two classifier algorithm functions of LibSVM, for further work or behavior to be carried out. Currently, people's living standards and the daily consumption of food adversely affect their health, especially their daily lives That increases the number of kidney diseases per day in India. His health criteria relied on people's diets earlier, now kidney disease is not only limited to people with diabetes or hypertension but it has several causes. All these things result in chemical cereals, vegetables and fruits; this is our normal diet, and the outcome is not where we are on the kidney

## II. METHODOLOGY



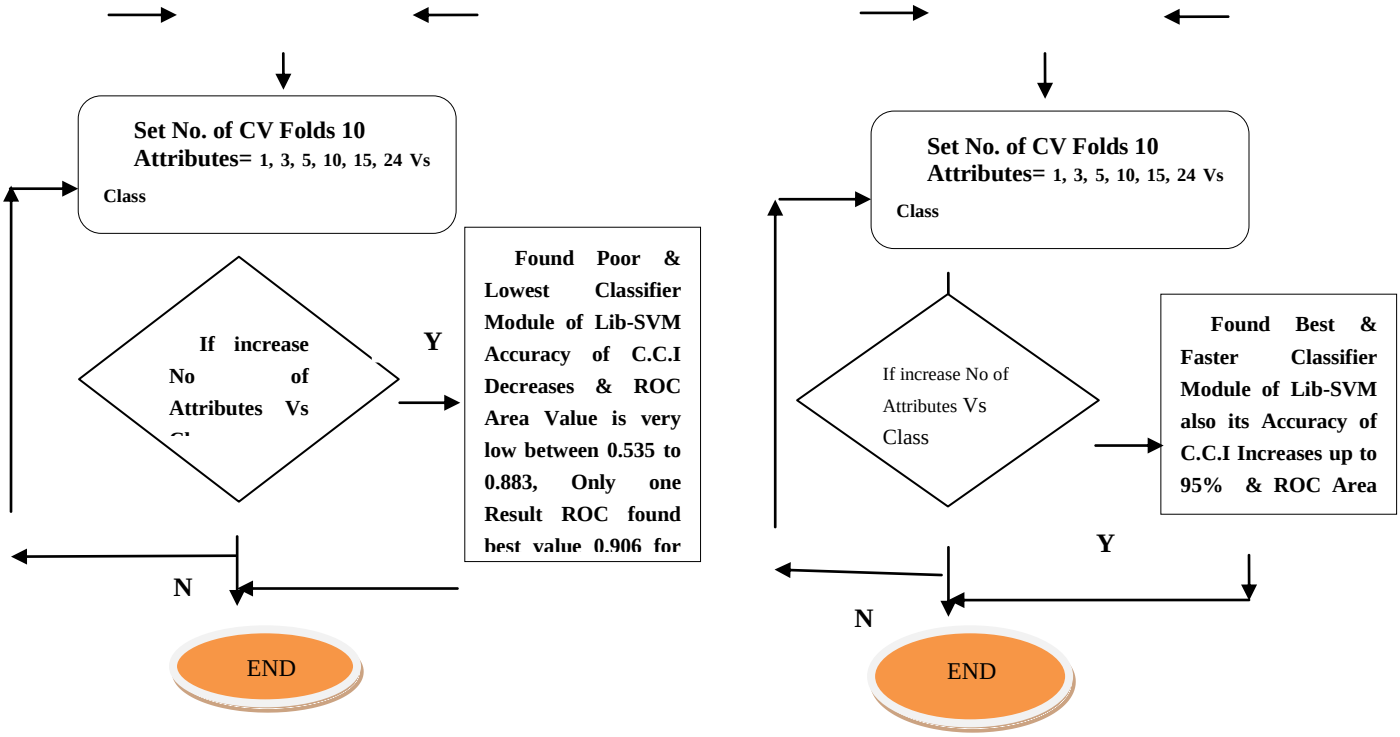



**Fig 1:** The research work methodology adopted

Usage of chronic kidney disease software, software training model perfection for LibSVM's Supervised Classification Learning Techniques – Support Vector Machine, and select some attribute parameters 1. RBC numbers 2. 3.Diabetes M: Hypertension (BP) 4. Cardiovascular disease 5.Appetite 6.The Edema 7 lever. Anemia, we use WEKA method to sort data using LibSVM algorithms. The clinical data of 400 reports of kidney disease eligible for review have been taken from the regular database for Machine Learning. The data collected for further review after cleaning and deleting missing values, the data comprises 25 attributes in the class dataset (CKD and Not-CKD) and Class distribution is (63 per cent for CKD and 37 per cent for non-CKD). Linear function efficiency is the highest in contrast with Polynomial function. Accordingly, the accuracy of the Correctly Categorized Instances (CCI) is 94.11% and ICI = 5.88% and the total number of instances approved by the program is 136, while the total number of attributes is 24, CVF=10 (Cross Validation Folds) and 66% and ROC = 0.945. Likewise, The highest accuracy found for CCI is = 95.00 percent and ICI=5.00 percent for attribute no uses 24, CVF=10 and 70 percent for splitting data and close to 1 i.e. 0.951 for the ROC values. All result will be shown and shown in Fig with the aid of visualizing classifier mistake. 3 Linear function, number of Attributes increases, module accuracy increases also in the form of CCI and ROC area values. The description of the classified model: contains the values of the correctly classified instances has been increased in terms of accuracy if the number of attributes has increased since 1,3,5,10,15 and 25, then the accuracy of the result increases almost to the ROC values 1.0 with the aid of the ROC values.

## III. RESULTS AND ANALYSIS

The outcome of the experiment is to be contrasted with Lib SVM Classifier's linear and polynomial function on the basis of accuracy in terms of high precision with a minimum processing time. Analyzing by data is the following algorithm; the performance and application of all three algorithms are as follows: Based on both test results with the aid of CKD dataset and Lib-SVM's Found Best &

Faster Classifier Module, its accuracy of C.C.I also increases by up to 95 percent and ROC Area by nearly 1.000 i.e. 0.951.

**3. 1. Linear function:**

Fig:1 followed technique, initially applying the Linear feature dataset, initially using one attribute and at the same time using 3,5,10,15 and 25 attributes to the 400 instances and finding the results. We use 1 attributes with Cross Validation Folds is 10 with 66 percent of the trained data set and the following results are found, and similarly for CVF is 10 with 70 percent of trained data set specific results found. The outcomes of the reports are presented in the outcome and discussion section in the form of table. The researcher would also apply the check for the attribute 5,10,15 and 25 on the same data with CVF=10 at 66 percent and 70 percent of the qualified data collection. Every result provides a description of the model check with correctly classified instances and incorrectly classified instances, on the basis of which the accuracy of the test model and the quantitative accuracy of the class table provides the TP Rate, FP Rate, Precision Value, Reminder, F-Measure, MCC, ROC Area, PRC Area and output of CKD, not CKD and Weighted Average Class values and fines.

**3. 2 Linear : --Visualize Classifier Error: for attributes 1,3,5,10,15 and 25 :➔**

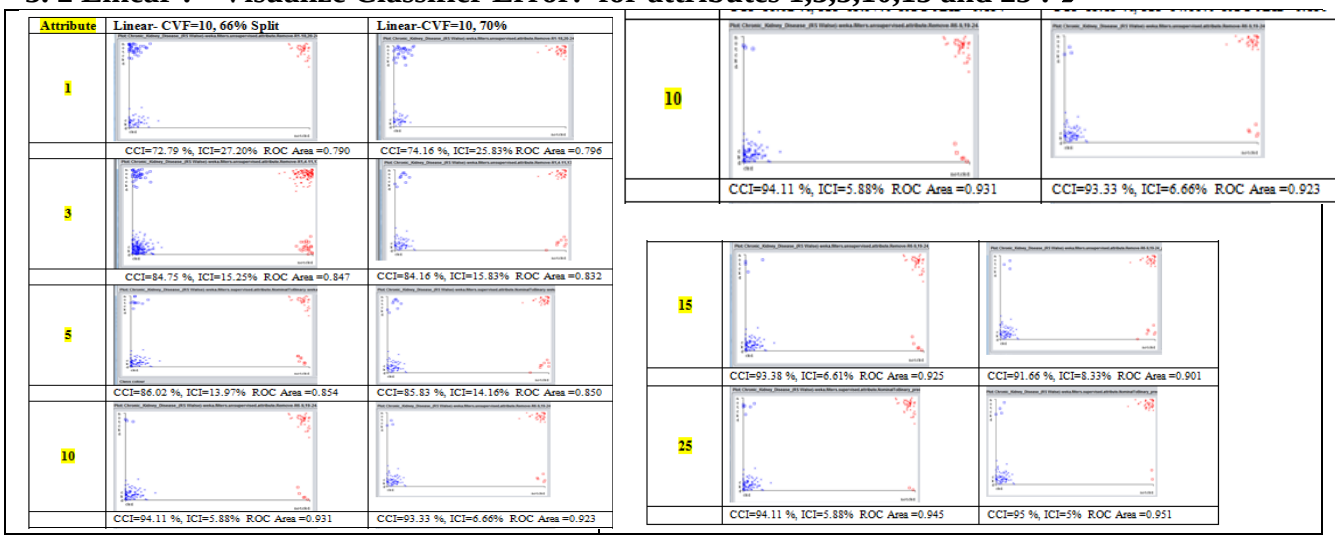



**Fig 2**: Linear function Visualization effect , CVF=10, 66% & 70%, attributes 1,3,5,10,15 & 25

**3.3 Used 1 attributes CVF=10, 70% trained data set**

=== Run information ===

Scheme: weka.classifiers.functions.LibSVM -S 0 -K 1 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -model "C:\\Program Files\\Weka-3-8-4" -seed 1

Relation: Chronic_Kidney_Disease_(RS Walse)-weka.filters.supervised.attribute.Remove-R1-18,20-24

Instances: 400

Attributes: 2

Htn, class

Test mode: split 70.0% train, remainder test

=== Classifier model (full training set) ===

LibSVM wrapper, original code by Yasser EL-Manzalawy (= WLSVM)

Time taken to build model: 0.02 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

| | | |
|---|---|---|
| Correctly Classified Instances | 89 | 74.1667 % |
| Incorrectly Classified Instances | 31 | 25.8333 % |
| Kappa statistic | 0.5156 | |
| Mean absolute error | 0.2583 | |
| Root mean squared error | 0.5083 | |
| Relative absolute error | 55.2173 % | |
| Root relative squared error | 105.4353 % | |
| Total Number of Instances | 120 | |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---|---|---|---|---|---|---|---|---|
| 0.592 | 0.000 | 1.000 | 0.592 | 0.744 | 0.589 | 0.796 | 0.850 | ckd |
| 1.000 | 0.408 | 0.587 | 1.000 | 0.739 | 0.589 | 0.796 | 0.587 | notckd |
| 0.742 | 0.150 | 0.848 | 0.742 | 0.742 | 0.589 | 0.796 | 0.754 | Weighted Av |

=== Confusion Matrix ===

| a | b | <-- classified as |
|---|---|---|
| 45 | 31 | a = ckd |
| 0 | 44 | b = notckd |

**3.4 Polynomial: --Visualize Classifier Error: for attributes 1,3,5,10,15 and 25 :➔**

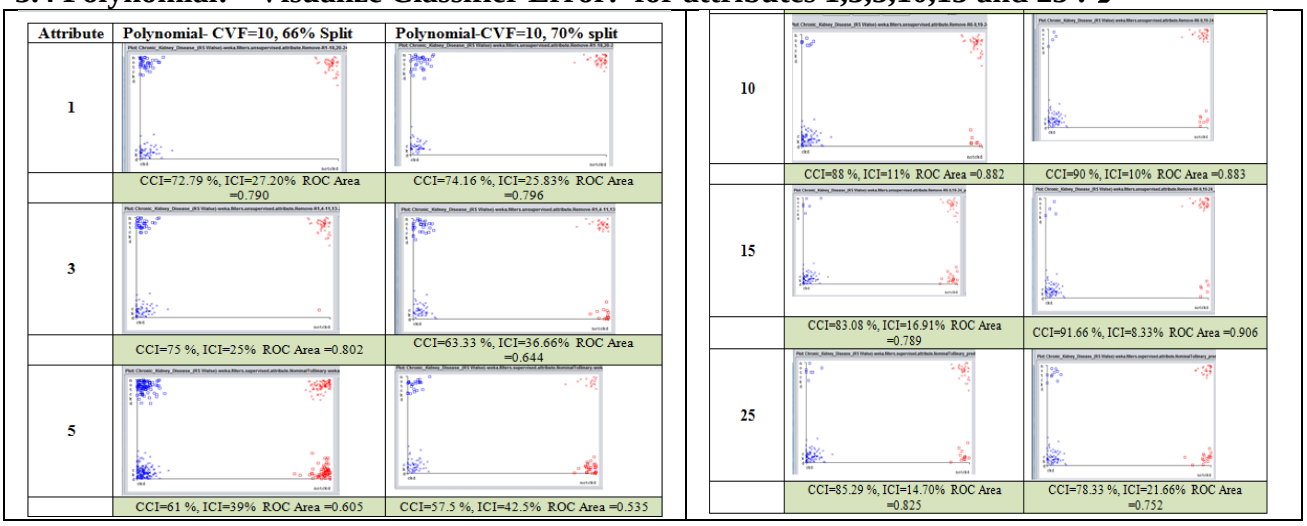



**Fig 3:** Polynomial function Visualization effect , CVF=10, 66% & 70%, attributes 1,3,5,10,15 & 25

## IV. RESULTS AND DISCUSSIONS

Columns are intended by analysis, the kidney data set processed with specific attributes (25), which comprises 400 rows, i.e. instances and 25 attributes. The researcher has selected each attribute to show the type of attributes, the form means Minimal, the number of missing values for each attribute in the data set. Instances, how many distinct values are present in the dataset, separate means different values, if we choose attribute Name- means nominal type in front of attribute-name. The statistical data in the form of descriptive statistics provides a description of the general results. Also, if data is qualitative then it is viewed as a class of attributes and displays its weight in the form of true / false or yes / no in the form of mark count.

**{a} Supervised : LibSVM- Classification – Algorithm using Polynomial and Linear function**

**4.1 Polynomial Function:**

Supervised- classify –LibSVM- Polynomial and Linear function – CVF=10 and split 66% and 70%

**Polynomial, CVF=10 , Split=66% and 70 % (Attributes= 1, 3, 5, 10, 15, 24)**

**Table 2:** Polynomial Correctness by Class values

| CVF / Split | No of Attributes | Class | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area |
|---|---|---|---|---|---|---|---|---|---|---|
| CVF=10 Split=66% | 01 | CKD | 0.580 | 0.000 | 1.000 | 0.580 | 0.734 | 0.572 | **0.790** | 0.852 |
| | | Not-CKD | 1.000 | 0.420 | 0.565 | 1.000 | 0.722 | 0.572 | 0.790 | 0.565 |
| | | Weight Avg | 0.728 | 0.148 | 0.846 | 0.728 | 0.730 | 0.572 | 0.790 | 0.750 |
| CVF=10 Split=70% | | CKD | 0.592 | 0.000 | 1.000 | 0.592 | 0.744 | 0.589 | **0.796** | 0.850 |
| | | Not-CKD | 1.000 | 0.408 | 0.587 | 1.000 | 0.739 | 0.589 | 0.796 | 0.587 |
| | | Weight Avg | 0.742 | 0.150 | 0.848 | 0.742 | 0.742 | 0.589 | 0.796 | 0.754 |
| CVF=10 Split=66% | 03 | CKD | 0.625 | 0.021 | 0982 | 0.625 | 0.764 | 0.587 | **0.802** | 0.856 |
| | | Not-CKD | 0.979 | 0.375 | 0.588 | 0.979 | 0.734 | 0.587 | 0.802 | 0.583 |
| | | Weight Avg | 0.750 | 0.146 | 0.843 | 0.750 | 0.753 | 0.587 | 0.802 | 0.760 |
| CVF=10 Split=70% | | CKD | 0.605 | 0.318 | 0.767 | 0.605 | 0.676 | 0.277 | 0.644 | 0.714 |
| | | Not-CKD | 0.682 | 0.395 | 0.500 | 0.682 | 0.577 | 0.277 | 0.644 | 0.458 |
| | | Weight Avg | 0.633 | 0.346 | 0.669 | 0.633 | 0.640 | 0.277 | 0.644 | 0.620 |
| CVF=10 Split=66% | 05 | CKD | 0.624 | 0.413 | 0.716 | 0.624 | 0.667 | 0.205 | 0.605 | 0.682 |
| | | Not-CKD | 0.587 | 0.376 | 0.484 | 0.857 | 0.530 | 0.205 | 0.605 | 0.439 |
| | | Weight Avg | 0.610 | 0.399 | 0.629 | 0.610 | 0.615 | 0.205 | 0.605 | 0.590 |
| CVF=10 Split=70% | | CKD | 0.684 | 0.614 | 0.658 | 0.684 | 0.671 | 0.072 | 0.535 | 0.650 |
| | | Not-CKD | 0.386 | 0.316 | 0.415 | 0.386 | 0.400 | 0.072 | 0.535 | 0.385 |
| | | Weight Avg | 0.575 | 0.504 | 0.569 | 0.575 | 0.572 | 0.072 | 0.535 | 0.553 |
| CVF=10 Split=66% | 10 | CKD | 0.909 | 0.146 | 0.920 | 0.909 | 0.914 | 0.760 | 0.882 | 0.895 |
| | | Not-CKD | 0.854 | 0.091 | 0.837 | 0.854 | 0.845 | 0.760 | 0.882 | 0.766 |
| | | Weight Avg | 0.890 | 0.126 | 0.890 | 0.890 | 0.890 | 0.760 | 0.882 | 0.849 |
| CVF=10 | | CKD | 0.947 | 0.182 | 0.900 | 0.947 | 0.923 | 0.783 | 0.883 | 0.886 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Split=70% | | Not-CKD | 0.818 | 0.053 | 0.900 | 0.818 | 0.857 | 0.783 | 0.883 | 0.803 |
| | | Weight Avg | 0.900 | 0.134 | 0.900 | 0.900 | 0.899 | 0.783 | 0.883 | 0.856 |
| CVF=10 Split=66% | **15** | CKD | 0.932 | 0.354 | 0.828 | 0.932 | 0.877 | 0.620 | 0.789 | 0.816 |
| | | Not-CKD | 0.646 | 0.068 | 0.838 | 0.644 | 0.729 | 0.620 | 0.789 | 0.666 |
| | | Weight Avg | 0.831 | 0.253 | 0.832 | 0.831 | 0.825 | 0.620 | 0.789 | 0.763 |
| CVF=10 Split=70% | | CKD | 0.947 | 0.136 | 0.923 | 0.947 | 0.935 | 0.819 | 0.906 | 0.908 |
| | | Not-CKD | 0.864 | 0.053 | 0.905 | 0.864 | 0.884 | 0.819 | 0.906 | 0.831 |
| | | Weight Avg | 0.917 | 0.106 | 0.916 | 0.917 | 0.916 | 0.819 | 0.906 | 0.880 |
| CVF=10 Split=66% | **24** | CKD | 0.920 | 0.271 | 0.862 | 0.920 | 0.890 | 0.672 | 0.825 | 0.845 |
| | | Not-CKD | 0.729 | 0.080 | 0.833 | 0.729 | 0.778 | 0.672 | 0.825 | 0.703 |
| | | Weight Avg | 0.853 | 0.203 | 0.852 | 0.853 | 0.850 | 0.672 | 0.825 | 0.795 |
| CVF=10 Split=70% | | CKD | 0.868 | 0.364 | 0.805 | 0.868 | 0.835 | 0.523 | 0.752 | 0.782 |
| | | Not-CKD | 0.636 | 0.132 | 0.737 | 0.636 | 0.683 | 0.523 | 0.752 | 0.602 |
| | | Weight Avg | 0.783 | 0.279 | 0.780 | 0.783 | 0.780 | 0.523 | 0.752 | 0.716 |

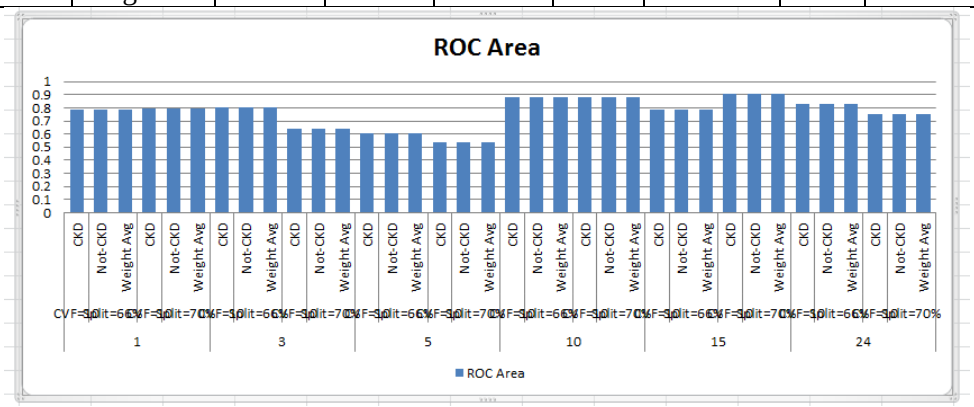



**Fig 4**: Shows the Graph of Polynomial Correctness by class values

**4.2 : Scheme:** weka.classifiers.functions.LibSVM -S 0 -K 1 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -model "C:\\Program Files\\Weka-3-8-4" -seed 1 **Relation:** Chronic_Kidney_Disease_(RS Walse)-weka.filters. supervised.attribute.Remove-R1-18,20-24 -LibSVM- Polynomial

**Table 3: Polynomial** Summary of Classifier model (Train set data)

| **Sr. No.** | **Particulars** | **No. of Attribute= 1** | | **No. of Attribute= 3** | | **No. of Attribute= 5** | | **No. of Attribute=10** | | **No. of Attribute=15** | | **No. of Attribute=24** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | TMSTRT | CVF =10 Split-66% | CVF =10 Split-70% | CVF =10 Split-66% | CVF =10 Split-70% | CVF =10 Split-66% | CVF =10 Split-70% | CVF =10 Split-66% | CVF =10 Split-70% | CVF =10 Split-66% | CVF =10 Split-70% | CVF =10 Split-66% | CVF =10 Split-70% |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **2** | TTBM | 0.02 | 0.02 | 11.92 | 11.72 | 6.44 | 5.69 | 4.55 | 4.38 | 3.71 | 3.31 | 5.0 | 4.4 |
| **3** | TTTM | 0 | 0 | 0 | 0 | - | 0 | 0 | 0 | 0.01 | 0.01 | 0.03 | 0 |
| **4** | CCI | **72.79** | **74.16** | **75** | **63.33** | **61** | **57.5** | **88.97** | **90** | **83.08** | **91.66** | **85.29** | **78.33** |
| **5** | ICI | 27.20 | 25.83 | 25 | 36.66 | 39 | 42.5 | 11.02 | 10 | 16.91 | 8.33 | 14.70 | 21.66 |
| | | 59 | | | | | | | | | | | |
| **6** | KS | 0.493 | 0.51 | 0.52 | 0.26 | 0.20 | 0.07 | 0.75 | 0.78 | 0.60 | 0.81 | 0.66 | 0.51 |
| | | 2 | | | | | | | | | | | |
| **7** | MAE | 0.272 | 0.25 | 0.25 | 0.36 | 0.39 | 0.42 | 0.11 | 0.1 | 0.16 | 0.08 | 0.14 | 0.21 |
| | | 1 | | | | | | | | | | | |
| **8** | RMSE | 0.521 | 0.50 | 0.5 | 0.60 | 0.62 | 0.65 | 0.33 | 0.31 | 0.41 | 0.28 | 0.38 | 0.46 |
| | | 6 | | | | | | | | | | | |
| **9** | RAE | 58.27 | 55.21 | 53.55 | 78.37 | 83.16 | 90.84 | 23.62 | 21.37 | 36.22 | 17.81 | 31.50 | 46.31 |
| | | 81 | | | | | | | | | | | |
| **10** | RRSE | 108.8 | 105.4 | 104.3 | 126.6 | 128.9 | 135.2 | 69.31 | 65.59 | 85.83 | 59.88 | 80.04 | 96.55 |
| | | 6 | 3 | 5 | 1 | 9 | 3 | | | | | | |
| **11** | TNI | 13 | 120 | 136 | 120 | 400 | 120 | 136 | 120 | 136 | 120 | 136 | 120 |
| | | 6 | 89 | 102 | 76 | 244 | 69 | 121 | 108 | 113 | 110 | 116 | 94 |
| | | 99 | 31 | 34 | 44 | 156 | 51 | 15 | 12 | 23 | 10 | 20 | 26 |
| | | 37 | | | | | | | | | | | |

| | | | |
|---|---|---|---|
| Test mode: split train, remainder test | TMSTRT | Mean Absolute Error | MAE |
| Time Taken to Build Model: | TTBM | Root Mean Squared Error | RMSE |
| Time Taken to Test Model on test split | TTTM | Relative Absolute Error | RAE |
| Correctly Classified Instances | CCI | Root Relative Squared Error | RRSE |
| Incorrectly Classified Instances | ICI | Total Number of Instances | TNI |
| Kappa statistic | KS | | |

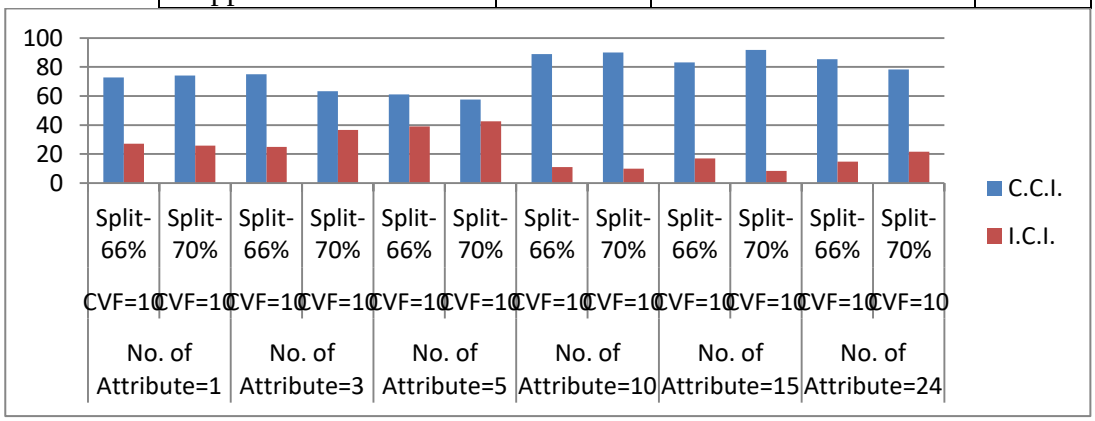



Fig: 5 Polynomial : Summary of Classified Model

**4.3 Linear , CVF=10 , Split=66% and 70 % (Attributes= 1, 3, 5, 10, 15, 24)**

**Table 4:** Linear Correctness by Class values

| **CVF/ Split** | **No of Attributes** | **Class** | **TP Rate** | **FP Rate** | **Precision** | **Recall** | **F-Measure** | **MCC** | **ROC Area** | **PRC Area** |
|---|---|---|---|---|---|---|---|---|---|---|
| CVF=10 Split=66 % | **01** | CKD | 0.580 | 0.000 | 1.000 | 0.580 | 0.734 | 0.572 | **0.790** | 0.852 |
| | | Not-CKD | 1.000 | 0.420 | 0.565 | 1.000 | 0.722 | 0.572 | 0.790 | 0.565 |
| | | Weight Avg | 0.728 | 0.148 | 0.846 | 0.728 | 0.730 | 0.572 | 0.790 | 0.750 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| CVF=10 Split=70 % | | CKD | 0.592 | 0.000 | 1.000 | 0.592 | 0.744 | 0.589 | **0.796** | 0.850 |
| | | Not-CKD | 1.000 | 0.408 | 0.587 | 1.000 | 0.739 | 0.589 | 0.796 | 0.587 |
| | | Weight Avg | 0.742 | 0.150 | 0.848 | 0.742 | 0.742 | 0.589 | 0.796 | 0.754 |
| CVF=10 Split=66 % | **03** | CKD | 0.848 | 0.153 | 0.902 | 0.848 | 0.874 | 0.683 | **0.847** | 0.860 |
| | | Not-CKD | 0.847 | 0.152 | 0.770 | 0.847 | 0.806 | 0.683 | 0.847 | 0.709 |
| | | Weight Avg | 0.848 | 0.153 | 0.852 | 0.848 | 0.849 | 0.683 | 0.847 | 0.803 |
| CVF=10 Split=70 % | | CKD | 0.868 | 0.205 | 0.880 | 0.868 | 0.874 | 0.661 | 0.832 | 0.848 |
| | | Not-CKD | 0.795 | 0.132 | 0.778 | 0.795 | 0.787 | 0.661 | 0.832 | 0.694 |
| | | Weight Avg | 0.842 | 0.178 | 0.843 | 0.842 | 0.842 | 0.661 | 0.832 | 0.791 |
| CVF=10 Split=66 % | **05** | CKD | 0.875 | 0.167 | 0.906 | 0.875 | 0.890 | 0.699 | 0.854 | 0.874 |
| | | Not-CKD | 0.833 | 0.125 | 0.784 | 0.833 | 0.808 | 0.699 | 0.854 | 0.712 |
| | | Weight Avg | 0.860 | 0.152 | 0.863 | 0.860 | 0.861 | 0.699 | 0.854 | 0.817 |
| CVF=10 Split=70 % | | CKD | 0.882 | 0.182 | 0.893 | 0.882 | 0.887 | 0.697 | 0.850 | 0.863 |
| | | Not-CKD | 0.818 | 0.118 | 0.800 | 0.818 | 0.809 | 0.697 | 0.850 | 0.721 |
| | | Weight Avg | 0.858 | 0.159 | 0.859 | 0.858 | 0.859 | 0.697 | 0.850 | 0.811 |
| CVF=10 Split=66 % | **10** | CKD | 0.966 | 0.104 | 0.944 | 0.966 | 0.955 | 0.870 | 0.931 | 0.934 |
| | | Not-CKD | 0.896 | 0.034 | 0.935 | 0.896 | 0.915 | 0.870 | 0.931 | 0.874 |
| | | Weight Avg | 0.941 | 0.079 | 0.941 | 0.941 | 0.941 | 0.870 | 0.931 | 0.913 |
| CVF=10 Split=70 % | | CKD | 0.961 | 0.114 | 0.936 | 0.961 | 0.948 | 0.856 | 0.923 | 0.924 |
| | | Not-CKD | 0.886 | 0.039 | 0.929 | 0.886 | 0.907 | 0.856 | 0.923 | 0.865 |
| | | Weight Avg | 0.933 | 0.086 | 0.933 | 0.933 | 0.933 | 0.856 | 0.923 | 0.902 |
| CVF=10 Split=66 % | **15** | CKD | 0.955 | 0.104 | 0.944 | 0.955 | 0.949 | 0.855 | 0.925 | 0.930 |
| | | Not-CKD | 0.896 | 0.045 | 0.915 | 0.896 | 0.905 | 0.855 | 0.925 | 0.856 |
| | | Weight Avg | 0.934 | 0.083 | 0.934 | 0.934 | 0.934 | 0.855 | 0.925 | 0.904 |
| CVF=10 Split=70 % | | CKD | 0.961 | 0.159 | 0.913 | 0.961 | 0.936 | 0.819 | 0.901 | 0.901 |
| | | Not-CKD | 0.841 | 0.39 | 0.925 | 0.841 | 0.881 | 0.819 | 0.901 | 0.836 |
| | | Weight Avg | 0.917 | 0.115 | 0.917 | 0.917 | 0.916 | 0.819 | 0.901 | 0.878 |
| CVF=10 Split=66 | **24** | CKD | 0.932 | 0.042 | 0.976 | 0.932 | 0.953 | 0.875 | 0.945 | 0.954 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| % | | Not-CKD | 0.958 | 0.068 | 0.885 | 0.958 | 0.920 | 0.875 | 0.945 | 0.862 |
| | | Weight Avg | 0.941 | 0.051 | 0.944 | 0.941 | 0.942 | 0.875 | 0.945 | 0.922 |
| CVF=10 Split=70 % | | CKD | 0.947 | 0.045 | 0.973 | 0.947 | 0.960 | 0.894 | 0.951 | 0.955 |
| | | Not-CKD | 0.955 | 0.053 | 0.913 | 0.955 | 0.933 | 0.894 | 0.951 | 0.888 |
| | | Weight Avg | 0.950 | 0.048 | 0.951 | 0.950 | 0.950 | 0.894 | 0.951 | 0.931 |

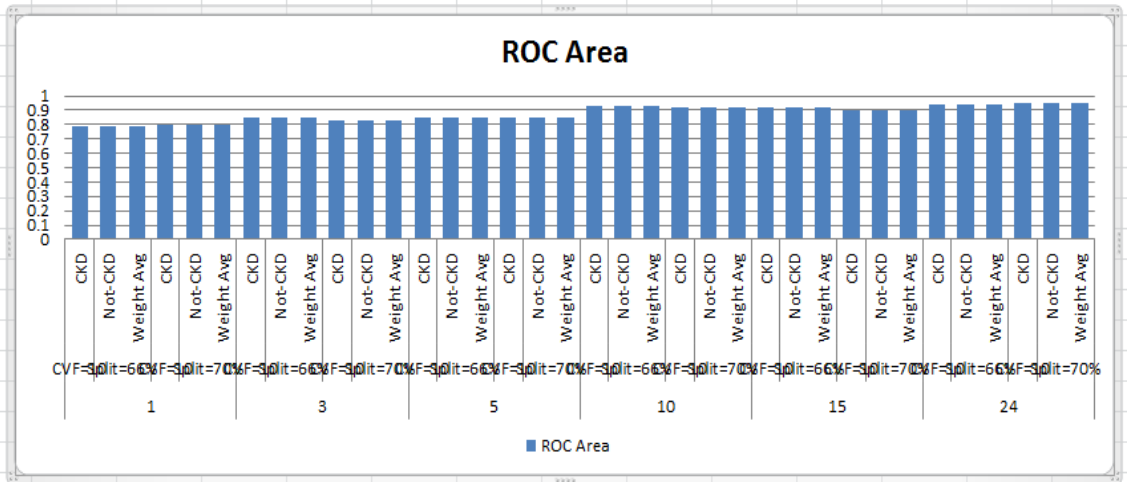



**Fig 6 :** Linear: Correctness by Class values

**4.4 Scheme:** weka.classifiers.functions.LibSVM -S 0 -K 0 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -model "C:\\Program Files\\Weka-3-8-4" -seed 1 **Relation:** Chronic_Kidney_Disease_(RS Walse)-weka.filters.supervised.attribute.Remove-R1-18,20-24

**Table 5:** LibSVM-Linear Summary of Classifier model (Train set data)

| **Sr. No.** | **Particulars** | **No. of Attribute=1** | | **No. of Attribute=3** | | **No. of Attribute=5** | | **No. of Attribute=10** | | **No. of Attribute=15** | | **No. of Attribute=24** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | TMSTRT | CVF=10 Split-66% | CVF=10 Split-70 % | CVF=10 Split-66 % | CVF=10 Split-70 % | CVF=10 Split-66 % | CVF=10 Split-70 % | CVF=10 Split-66 % | CVF=10 Split-70 % | CVF=10 Split-66 % | CVF=10 Split-70 % | CVF=10 Split-66 % | CVF=10 Split-70 % |
| **2** | TTBM | 0 | 0.02 | 0.16 | 0.02 | 0.81 | 0.14 | 0.17 | 0.16 | 2.74 | 2.37 | 3.21 | 2.82 |
| **3** | TTTM | 0 | 0 | - | 0 | 0 | 0 | 0 | 0 | 0.01 | 0.02 | 0.01 | 0.01 |
| **4** | CCI | **72.79** | **74.16** | **84.75** | **84.16** | **86.02** | **85.83** | **94.11** | **93.33** | **93.38** | **91.66** | **94.11** | **95.00** |
| **5** | ICI | 27.20 | 25.83 | 15.25 | 15.83 | 13.97 | 14.16 | 5.88 | 6.66 | 6.16 | 8.33 | 5.88 | 5.00 |
| **6** | KS | 0.49 | 0.51 | 0.61 | 0.66 | 0.69 | 0.69 | 0.87 | 0.85 | 0.85 | 0.81 | 0.87 | 0.89 |
| **7** | MAE | 0.27 | 0.25 | 0.15 | 0.15 | 0.13 | 0.14 | 0.05 | 0.06 | 0.06 | 0.08 | 0.05 | 0.05 |
| **8** | RMS | 0.52 | 0.50 | 0.39 | 0.39 | 0.37 | 0.37 | 0.24 | 0.25 | 0.25 | 0.28 | 0.24 | 0.22 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | E | | | | | | | | | | | | |
| **9** | RAE | 58.27 | 55.21 | 32.52 | 33.84 | 29.92 | 30.28 | 12.60 | 14.24 | 14.17 | 17.81 | 12.60 | 10.68 |
| **10** | RRSE | 108.86 | 105.43 | 80.66 | 82.54 | 78.01 | 78.07 | 50.62 | 53.56 | 53.69 | 59.88 | 50.62 | 46.38 |
| **11** | TNI | 1 36 9 9 3 7 | 120 89 31 | 400 339 61 | 120 101 19 | 136 117 19 | 120 103 17 | 136 128 8 | 120 112 8 | 136 127 9 | 120 110 10 | 136 128 8 | 120 114 6 |

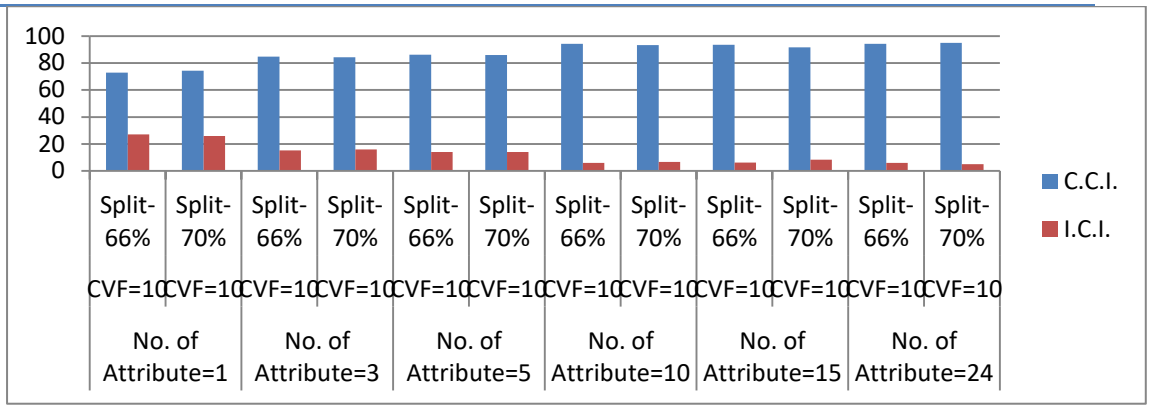



**Fig 7** : Linear: Summary of Classified Model

**4.5 Confusion Matrix:**

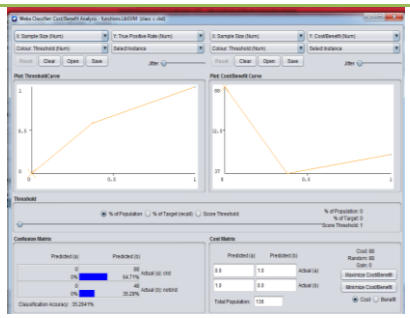
How to assess our model 's effectiveness, improve the quality, improve the results and exactly what we want. So it is the matrix of Confusion that falls into the limelight. Confusion Matrix is a performance metric for classification of machine learning, where results may be two or more groups. This is extremely useful for calculating Recall, Precision, Specificity, Accuracy and the importance of ROC Curve Area most notably. The outcome of the confusion matrix was traced by applying linear and polynomial function with CVF=10, with 66 percent & 70 percent qualified data set with attributes ranging from 1,3,5,10,15 and 24. The product of the uncertainty matrix listed in terms of class values as CKD and not-CKD as the expected values of a and b as.

**Table 6**: Confusion Matrix : LibSVM Classifier – with Linear function

| **Sr. No.** | **No. of Attributes** | **CVF=10, Split- 66%** | | **CVF=10, Split- 70%** | | **< - Classified as** |
|---|---|---|---|---|---|---|
| | | Predicted (a) | Predicted (b) | Predicted (a) | Predicted (b) | |
| **1** | **1** | 51 | 37 | 45 | 31 | \| a = ckd |
| | | 0 | 48 | 0 | 44 | \| b = not-ckd |
| **2** | **3** | 212 | 38 | 66 | 10 | \| a = ckd |
| | | 23 | 127 | 9 | 35 | \| b = not-ckd |
| **3** | **5** | 77 | 11 | 67 | 9 | \| a = ckd |
| | | 8 | 40 | 8 | 36 | \| b = not-ckd |
| **4** | **10** | 85 | 3 | 73 | 3 | \| a = ckd |
| | | 5 | 43 | 5 | 39 | \| b = not-ckd |
| **5** | **15** | 84 | 4 | 73 | 3 | \| a = ckd |
| | | 5 | 43 | 7 | 37 | \| b = not-ckd |
| **6** | **24** | 82 | 6 | 72 | 4 | \| a = ckd |
| | | 2 | 46 | 2 | 42 | \| b = not-ckd |

**Table 7:** Confusion Matrix : LibSVM – Clasifier- with Polynomial function

| Sr. No. | No. of Attributes | *CVF=10, Split- 66% | | *CVF=10, Split- 70% | | < - Classified as |
|---|---|---|---|---|---|---|
| | | Predicted (a) | Predicted (b) | Predicted (a) | Predicted (b) | |
| **1** | **1** | 51 | 37 | 45 | 31 | \| a = ckd |
| | | 0 | 48 | 0 | 44 | \| b = not-ckd |
| **2** | **3** | 55 | 33 | 46 | 30 | \| a = ckd |
| | | 1 | 47 | 14 | 30 | \| b = not-ckd |
| **3** | **5** | 156 | 94 | 52 | 24 | \| a = ckd |
| | | 62 | 88 | 27 | 17 | \| b = not-ckd |
| **4** | **10** | 80 | 8 | 72 | 4 | \| a = ckd |
| | | 7 | 41 | 8 | 36 | \| b = not-ckd |
| **5** | **15** | 82 | 6 | 72 | 4 | \| a = ckd |
| | | 17 | 31 | 6 | 38 | \| b = not-ckd |
| **6** | **24** | 81 | 7 | 66 | 10 | \| a = ckd |
| | | 13 | 35 | 16 | 28 | \| b = not-ckd |



* CVF = Cross Validation Folds

Table 6 & 7 shows the visualization effect of linear feature uncertainty matrix, so we can easily predict and quantify the outcome using cost / benefit analysis based on the graphical output also shows the product of the uncertainty matrix, i.e. expected value of a and b with CKD class and not-CKD class.

It observed that the precision level of the Linear function of the LibSVM- classifier is best compared to the LibSVM classifier algorithm of polynomial function among the classification of CKD and Not-CKD patients. In addition, the ROC curve (Receiver Operating Characteristics Curve) area is close to 1.0 of Linear function allows patients with CKD and Not-CKD to be classified by Linear function , i.e. 95% accuracy. It is further found that, relative to the Polynomial LibSVM algorithm, it is more accurate. Figures 6 and 7 show the representation of the graph finding the result with the input value of CVF=10, with 66 percent and 70 percent of the qualified data set of CKD patients having attributes 1,3,5,10,15,20 and 24 of the linear and polynomial function. Figure 6 and 7 for attribute 2 with 66 percent and 70 percent trained and for polynomial for attribute 2 with 66 percent and 70 percent of trained data set and remaining output shows the result in table format in Table. 2,3,4 & 5.

## V. CONCLUSIONS

This study confirms that the LibSVM classifier algorithm of Linear model type directly impacts the treatment of CKD patients on the basis of prediction and accurate outcome of the Linear model. To offer care to patients with CKD. Such results suggest that tension can be minimized in the medical industry, and their adherence to treatment protocols may be strengthened by offering medical assessment training in the LibSVM Linear function model. Use of LibSVM Supervised algorithm for the classification. Linear function while increasing the number of attributes, the accuracy of the Linear model increased with respect to correctly classified instances and the value of receiver operating characteristic curve areas as well as the exact result of confusion matrix gain. The linear method of the LibSVM classifier algorithm is the best and provides the most reliable and correct

results for 66 percent and 70 percent of the split value (training on a portion of the data and checking on the rest) accuracy is 94 percent accuracy for 66 percent split trained data and 95 percent accuracy for 70 percent split trained data and 1 percent ROC value. The analysis limitation is comparative evaluation of two classifier algorithm functions of LibSVM, for further work or behavior to be carried out.

## VI. ACKNOWLEDGEMENT

The authors are grateful to the UCI ML repository for providing all types of WEKA database necessary to provide such a reliable tool for the extraction and analysis of knowledge from the database

## REFERENCES

1. Arasu, S. Dilli, and R. Thirumalaiselvi. "A novel imputation method for effective prediction of coronary Kidney disease." In *2017 2nd International Conference on Computing and Communications Technologies (ICCCT)*, pp. 127-136. IEEE, 2017.
2. Ariff, M. H., I. Ismarani, and N. Shamsuddin. "RFID based systematic livestock health management system." In *2014 IEEE Conference on Systems, Process and Control (ICSPC 2014)*, pp. 111-116. IEEE, 2014.
3. Bharara, Sanyam, A. SaiSabitha, and AbhayBansal. "A review on knowledge extraction for business operations using data mining." In *2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence*, pp. 512-518. IEEE, 2017.
4. Chuan, Zou, Tang Ying, Bai Li, ZengYuqun, and Lu Fuhua. "Application of clustering analysis to explore syndrome evolution law of peritoneal dialysis patients." In *2013 IEEE International Conference on Bioinformatics and Biomedicine*, pp. 23-26. IEEE, 2013.
5. Due ThanhAnhLuong, Dept of the computer. Sci&Engi, uni, at Buffalo Buffalo, NY, USAA K-Means Approach to Clustering disease Progressions IEEE Keywords, Sep 2017
6. Güllüoğlu, SabriSerkan. "Segmenting customers with data mining techniques." In *2015 Third International Conference on Digital Information, Networking, and Wireless Communications (DINWC)*, pp. 154-159. IEEE, 2015.
7. Jinyin, Chen, et al. "A novel cluster center fast determination clustering algorithm." *Applied Soft Computing* 57 (2017): 539-555.
8. Khanna, Umesh. "The economics of dialysis in India." *Indian journal of nephrology* 19, no. 1 (2009): 1.
9. Kunwar, Veenita, et al. "Chronic Kidney Disease analysis using data mining classification techniques." *2016 6th International Conference-Cloud System and Big Data Engineering (Confluence)*. IEEE, 2016.
10. Narander Kumar, SabitaKLhatri, Department of Computer, 3rd IEEE International Conference on Computational Intelligence and Communication Technology (IEEE-CICT 2017) " Implementing WEKA for medical data classification and early disease prediction 978-1-50, 2017
11. Uboltham, Issariya, NakornthipPrompoon, and Wirichada Pan-Ngum. "AKIHelper: Acute kidney injury diagnostic tool using KDIGO guideline approach." In *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, pp. 1-6. IEEE, 2016.
12. Khanna, Umesh. "The economics of dialysis in India." *Indian journal of nephrology* 19, no. 1 (2009): 1.
13. Varma, P. P. "Prevalence of chronic kidney disease in India-Where are we heading?." *Indian journal of nephrology* 25, no. 3 (2015): 133.
14. https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease