MULTI-DISEASE PREDICTION MODEL USING IMPROVED SVM-RADIAL BIAS TECHNIQUE IN HEALTHCARE MONITORING SYSTEM

Dr.M.Rajaiah[1],Dean Academics &HOD,Dept ofCSE,AudisankaraCollegeofsEngineeringandTechnology,Gudur.

Dr.U.Thirupalu[2], Associate Professor ,Dept of CSE, Audisankara CollegeofEngineeringandTechnology,Gudur.

Mr.Kakavakam Sandeep[2], UGScholar,DeptofCSE,AudisankaraCollegeofEngineeringandTechnology,Gudur.

Ms.Gogula Prameela[3],UGScholar, DeptofCSE,AudisankaraCollegeofEngineeringandTechnology,Gudur

Mr.BollineniJayanth Krishna[4],UGScholar, Dept ofCSE,AudisankaraCollege ofEngineeringandTechnology,Gudur.

Mr.Annam Sunil[5],UGScholar,DeptofCSE,AudisankaraCollegeofEngineeringandTechnology, Gudur

ABSTRACT_ Data is a valuable resource in this digital age, and there was a tonne of data being generated across all fields. Health care industry data comprises contains information about the patient as well as information on the ailment. The use of machine learning and medical data will assist us in doing data analysis to uncover hidden illness patterns and develop individualized treatments for the patient and utilized to forecast the illness. A generic architecture for illness prediction has been presented in this work in the field of medicine. Using an improved SVM-Radial bias kernel method, this system was tested using a smaller set of features from the Chronic Kidney Disease, Diabetes, and Heart Disease dataset. It was also compared to other machine learning methods such as SVM-

Linear, SVMPolynomial, Random forest, and Decision tree in R studio. All of these machine learning algorithms have been assessed based on their performance

1.INTRODUCTION

Due to digitalization, data have been rising exponentially in all fields in recent years. Massive data is referred to as "big data," and it cannot be handled by standard computers. Big data analytics is the process of analyzing huge databases to find hidden patterns, new information, and value. Numerous uses of big data analytics include weather forecasting, fraud and risk detection, logistical delivery, and healthcare .We may investigate algorithms that employ enormous data sets to learn, generalize, and forecast with the use of machine learning algorithms. Making judgments and computational statistics are both directly connected to machine learning**.**

Applications for machine learning algorithms include forecasting product sales, determining the likelihood that rain will fall in a specific area, and many more. Medical professionals can identify illness patterns and determine the severity of the condition by conducting a systematic study of the available medical data. Systematic analysis combined with machine learning algorithms will enable us to create prediction models for individualized therapy, keep track of patients warning signals while they undergo a trial, and help doctors choose the right medication for each patient In this study, we presented a decision support system integration framework for forecasting the illness using machine learning methods.

Improved SVM-Radial bias kernel methodology was used to create the decision support system for illness prediction, and its effectiveness was measured against that of other machine learning methods.

2.RELATED WORKS

Different researchers are doing research in clinical data analytics to predict various diseases by using diferent machine learning algorithms. This section discusses about various studies in health care analytics, chronic kidney diseases, diabetes, heart disease and diferent machine learning algorithms used in clinical data analytics.

In Li et al. (2019), the authors described the potential advantages of data analytics in the health care industry. The authors also described research opportunities in the healthcare industry to construct predictive modelling, statistical and algorithms to improve the clinical trial design. Data preprocessing will remove unwanted and noisy data from the dataset.

Misir et al. (2017) had reported that, the reduced set features for chronic kidney disease prediction using correlationbased algorithms. In this model, the authors considered only eight factors out of twenty-fve factors of the UCI kidney disease dataset to predict the diseases

The authors in Norouzi et al. (2016) constructed an integrated intelligent fuzzy expert system for predicting renal failure progression. In this model, they used 15 cc/kg/ min/1.73m2 as a GFR threshold value for predicting renal failure. Barakat et al. had proposed SVM based tool for predicting the diabetes disease. In that experiment, they recorded the prediction accuracy as 94%. With this motivation and literature survey, we extended to develop a Health care monitoring system and decision support system for predicting the disease using healthcare data

Shen et al. (2016) have proposed the fruit fly optimization to fnd out the appropriate parameters for improving the classifcation accuracy. In Polat et al., the authors used the SVM algorithm with Best First search engine feature selection method to predict chronic kidney diseases. For this experiments, the authors have

3.HEALTHCARE ANALYTICAL MODEL

The present work has proposed for predicting multiple diseases in the Health care industry. As a test phase, the system has experimented with Chronic Kidney Disease (CKD),Diabetes and Heart diseases from UCI dataset with the guidelines of the National Kidney Foundation, Standards of Medical Care in Diabetes, Cardiovascular Clinical Recommendations and Guidelines respectively. The main objective of the system is to help the doctors to reduce the diagnosis time so that doctors can start the treatment at earliest. The correctness of the model is computed using the confusion matrix. The confusion matrix is formulated with actual count and predicted count. The confusion matrix represented with the number of correct and incorrect predictions and it is represented in Table 1

TRUE POSITIVE Predicting the patient with the disease as yes, and its observation results are yes.
TRUE NEGATIVE Predicting the patient with no disease and its observation results are negative.
FALSE POSITIVE Predicting the patient with no diseases but with the observation, the patient is having the disease.
FALSE NEGATIVE Predicting the patient doesn`t have a disease but the patient has the disease.

ACTUAL	PREDICTOR	
	YES	NO
YES	TRUE POSITIVE	FALSE NEGATIVE
NO	FALSE POSITIVE	TRUE NEGATIVE

Table 1 Confusion Matrix

EVALUATION PARAMETERS:

A.ACCURACY:

The accuracy of the classifier id described as how many samples were corrected accurately among the total number of samples and it is represented as follows

$$\text{Accuracy} = \frac{\textbf{True Positive+True Negative}}{\textbf{Total no of Predictions}} * 100$$

B.MISCLASSIFICATION RATE:

It is described as, the classifier which is not predicting properly and it is also referred to as error rate

$$\text{Misclassification Rate} = \frac{\textbf{False Positive+False Negative}}{\textbf{Total no of Predictions}} * 100$$

C.RECALL:

It is described as the classifier which is predicting the positive values to actual positive values

$$\text{Recall} = \frac{\textbf{True Positive}}{\textbf{False Positive+False Negative}}$$

D. PRECISION:

The classifier which is predicting the positive values to the total positive values

$$\text{Precision} = \frac{\textbf{True Positive}}{\textbf{True Positive+False Positive}}$$

E.SPECIFICITY:

It is a measurement of predicting actual negatives as negative

$$\text{Specificity}=\frac{\textbf{True Negative}}{\textbf{True Negative+False Positive}}$$

4. RESULTS AND DISCUSSION

The system has experimented with CKD, Diabetes and Heart disease datasets from the UCI repository using R tool. During this experiment, the dataset was split into two parts as 80% for training and 20% for validation. We used Chi square method to identify the important features in each disease dataset.

From the reduced features identified as important features from feature selection phase, we implemented with SVM-Linear, SVMPolynomial, Random forest method, Decision tree method and Improved SVM- Radial method. Here SVM-Radial bias method was using variance reduction technique, with "C" and "gamma" values as 1 and 0.60 respectively. The tested results are evaluated with accuracy, misclassification rate, sensitivity, precision and specificity. Shows the prediction Accuracy, Misclassification rate, Specificity, Precision and sensitivity of CKD, Heart Disease and Diabetes dataset using various machine learning algorithms. From the results, we found that Improved SVM-Radial bias kernel technique was generating better results well while comparing to other methods in these three diseases

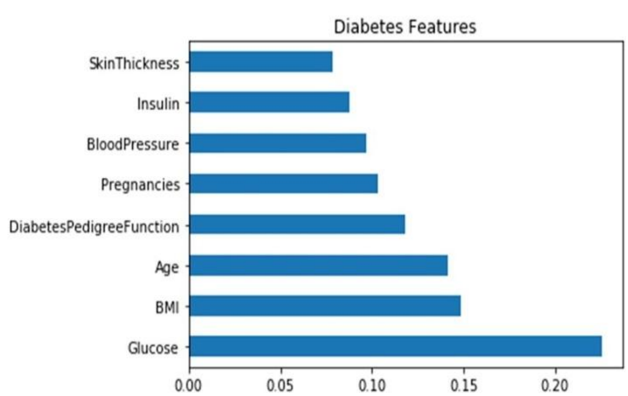


Fig. 4 Important Features In Diabetes

ACCURACY

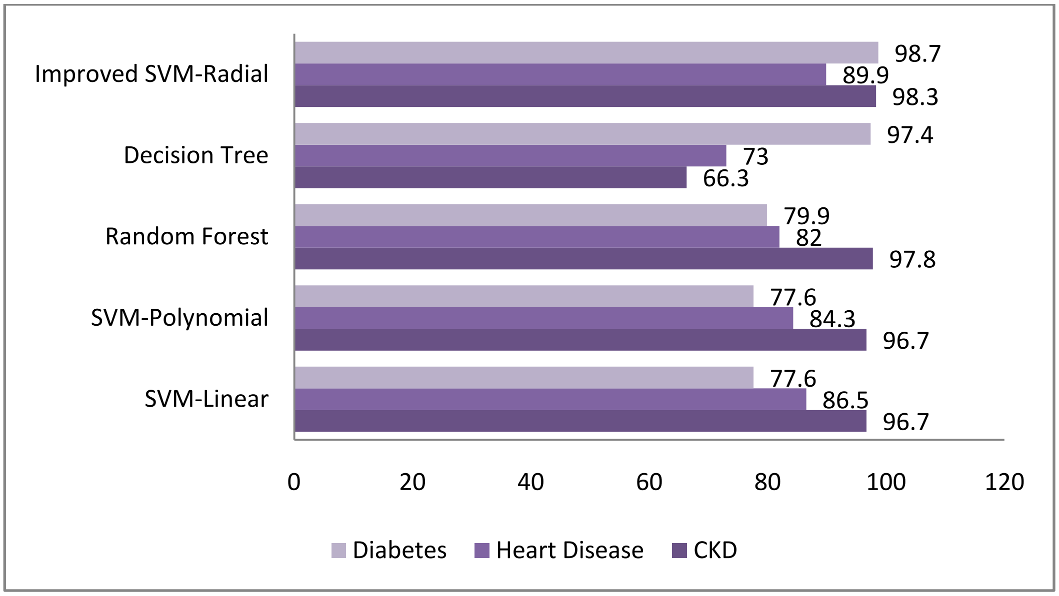


Fig.7 Comparison of Accuracy using Improved SVM-Radial Bias Kernel,decision tree random forest,SVM-Polynomial & SVM-Linear

In the Accuracy Graph The Improved SVM-Radial for Diabetes is 98.7%,Heart Disease is 89.9%,CKD is 98.3%.The Decision Tree for Diabetes is 97.4%,Heart Disease is 73.0%,CKD is 66.3%.The Random Forest for Diabetes is 79.9%,Heart Disease is 82.0%,CKD is 97.8%.The SVM-Polynomial for Diabetes is 77.6%,Heart Disease is 84.3%,CKD is 96.7%.The SVM-Linear for Diabetes is 77.6%,Heart Disease is 86.5%,CKD is 96.7%

MISCLASSIFICATION RATE:

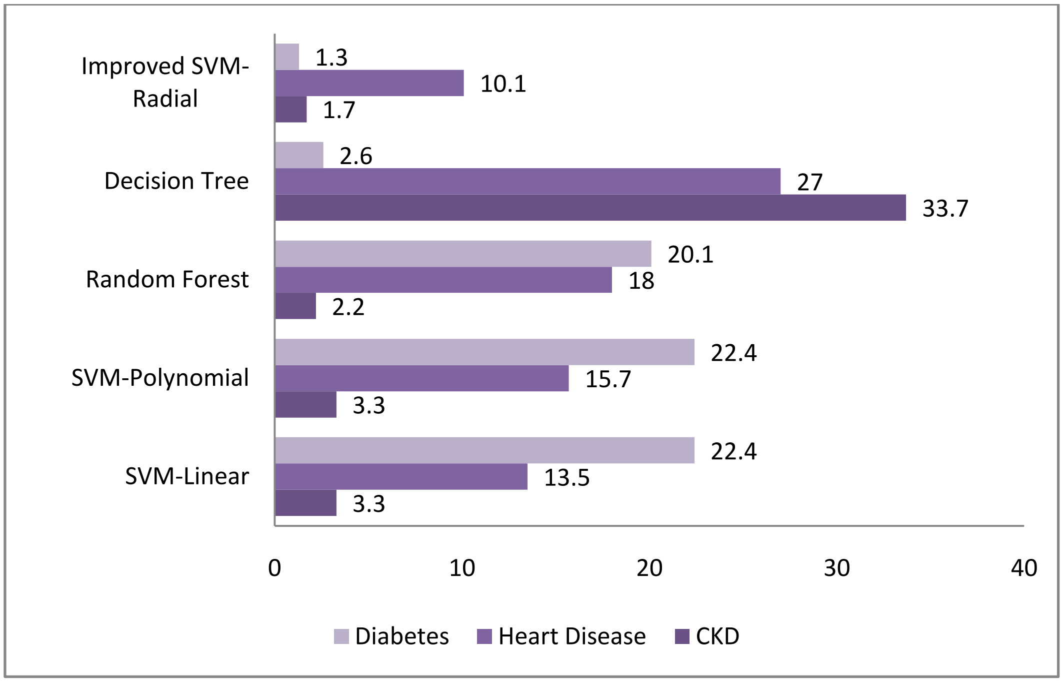


Fig.8 Comparison of Misclassification using Improved SVM-Radial Bias Kernel,decision tree random forest,SVM-Polynomial & SVM-Linear

In the Misclassification Rate Graph The Improved SVM-Radial for Diabetes is 1.3%,Heart Disease is 10.1%,CKD is 1.7%.The Decision Tree for Diabetes is 2.6%,Heart Disease is 27.0%,CKD is 33.7%.The Random Forest for Diabetes is 20.1%,Heart Disease is 18.0%,CKD is 2.2%.The SVM-Polynomial for Diabetes is 22.4%,Heart Disease is 15.7%,CKD is 3.3%.The SVM-Linear for Diabetes is 22.4%,Heart Disease is 13.5%,CKD is 3.3%

PRECISION:

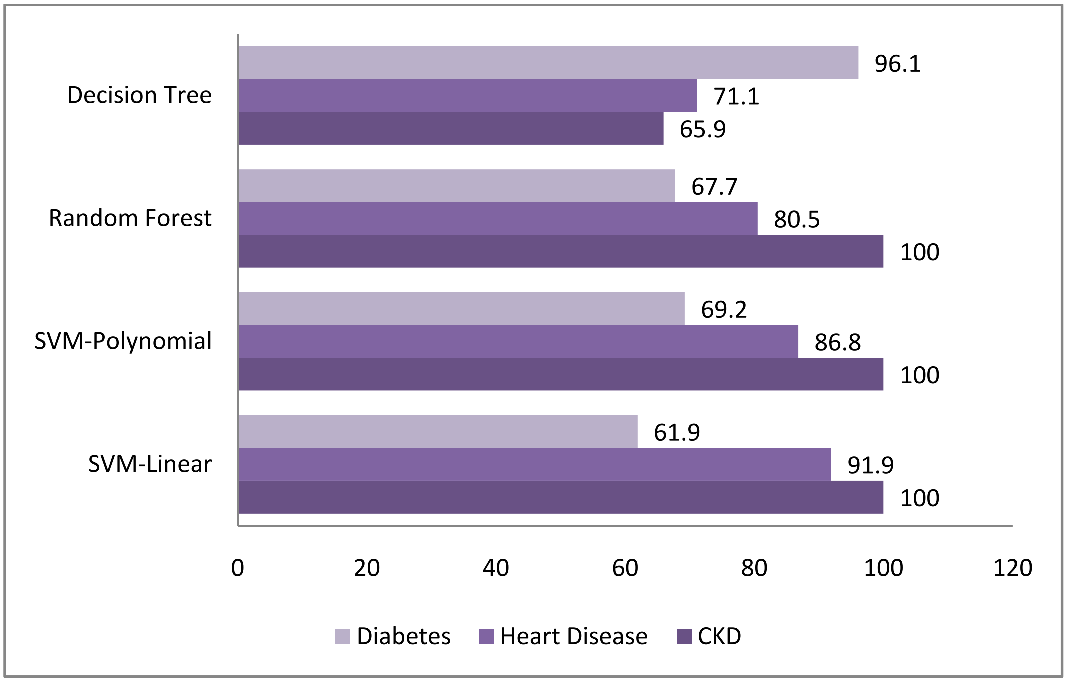


Fig.10 Comparison of Precision Graph using Improved SVM-Radial Bias Kernel,decision tree random forest,SVM-Polynomial & SVM-Linear

In the Precision Graph,The Decision Tree for Diabetes is 96.1%,Heart Disease is 71.1%,CKD is 65.9%.The Random Forest for Diabetes is 67.4%,Heart Disease is 80.5%,CKD is 100%.The SVM-Polynomial for Diabetes is 69.2%,Heart Disease is 86.8%,CKD is 100%.The SVM-Linear for Diabetes is 61.9%,Heart Disease is 91.9%,CKD is 100%

5.CONCLUSION:

In this work, we proposed a general architecture decision support system to predict multiple diseases in medical diagnosis. The system has experimented with CKD, Heart Disease and Diabetes disease datasets from UCI repository. For this experiment, essential features were identifed using Square method

Model		SVM-Linear (%)	SVM-Polynomial (%)	Random Forest (%)	Decision Tree (%)	Improved SVM-Radial (%)
Evaluation Parameters	**Disease**					
Accuracy	**CKD**	96.7	96.7	97.8	66.3	**98.3**
	Heart Disease	86.5	84.3	82.0	73.0	**89.9**
	Diabetes	77.6	77.6	79.9	97.4	**98.7**
Misclassification	**CKD**	3.3	3.3	2.2	33.7	**1.7**
	Heart Disease	13.5	15.7	18.0	27.0	**10.1**
	Diabetes	22.4	22.4	20.1	2.6	**1.3**
Sensitivity	**CKD**	100.0	100.0	100.0	65.9	**100.0**
	Heart Disease	91.9	86.8	80.5	71.1	**97.2**
	Diabetes	61.9	69.2	67.4	96.1	**95.5**
Precision	**CKD**	90.0	90.0	94.7	65.9	**95.0**
	Heart Disease	79.1	78.6	80.5	83.3	**81.4**
	Diabetes	59.1	40.9	66.0	96.1	**100.0**
Specificity	**CKD**	95.2	95.2	96.3	66.7	**97.6**
	Heart Disease	82.7	82.4	83.3	74.5	**84.9**
	Diabetes	**83.6**	**79.4**	**85.2**	**98.1**	**100.0**

Table 2 Analyzing The Prediction Accuracy And Misclassifcation Rate Of CKD, Heart Disease And Diabetes Using Machine Learning Algorithms

REFERENCES:

Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S (2016) Lung pattern classifcation for interstitial lung diseases using a deep convolutional neural network. IEEE Trans Med Imaging 35(5):1207–1216

Barakat N, Bradley AP, Barakat MNH (2010) Intelligible support vector machines for diagnosis of diabetes mellitus. IEEE Trans InfTechnol Biomed 14(4):1114–1120

Becker C, Gather U (2001) The largest nonidentifable outlier: a comparison of multivariate simultaneous outlier identifcation rules. Comput Stat Data Anal 36(1):119–127

Black N, Payne M (2003) Directory of clinical databases: improving and promoting their use. BMJ QualSaf 12(5):348–352

Burges CJ (1998) A tutorial on support vector machines for pattern recognition. Data Min Knowl Disc 2(2):121–167

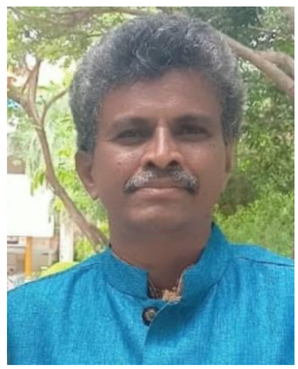
Chahal D, Gulia P (2016) Big data analytics. Res J ComputInfTechnolSci 4(2):1–4

Chen M, Hao Y, Hwang K, Wang L, Wang L (2017) Disease prediction by machine learning over big data from healthcare communities. IEEE Access 5:8869–8879

Çomak E, Arslan A, Türkoğlu İ (2007) A decision support system based on support vector machines for diagnosis of the heart valve diseases. ComputBiol Med 37(1):21–27

AUTHORPROFILES

Dr.M.Rajaiah,CurrentlyworkingasanDeanAcademics&HODinthedepartmentofCSEatASCET(Autonomous),Gudur,Tirupathi(DT).Hehaspublishedmorethan35papers in,WebofScience,ScopusIndexing,UGCJournals.



Dr.U.Thirupalu completed his Bachelor of Technology in Computer Scienceand Engineering.He completed his Masters of Technology in Computer Science andEngineering. Awarded Ph.D in Computer Science and Engineering at PondicherryUniversity (CentralUniversity), Puducherry. He haspublished more than12 papersin indexing Journals.Currently working as an Associate Professor in the departmentofCSEatASCET(Autonomous),Gudur,Tirupathi(DT).Hisareasofinterestinclude, Data Mining, CloudComputingand MachineLearning.

Mr. **Kakavakam Sandeep**, asB.Tech student in the department of CSE at

AudisankaraCollege of Engineering and Technology, Gudur

Ms. **GogulaPrameela**, asB.Tech student in the department of CSE at AudisankaraCollege of Engineering and Technology, Gudur

Mr. **BollineniJayanth Krishna**, asB.Tech student in the department of CSE at AudisankaraCollege of Engineering and Technology, Gudur

Mr. **Annam Sunil**, asB.Tech student in the department of CSE at AudisankaraCollege of Engineering and Technology, Gudur