ISSN 2347 - 3983
Volume 9. No. 7, July 2021
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter14972021.pdf
https://doi.org/10.30534/ijeter/2021/14972021

Comparative Study of Classification algorithms used for the Prediction of Non-communicable diseases

Veena Kumari H M[1], Dr. D S Suresh[2]
[1]Research Scholar,Department of ECE, Channabasaveshwara Institute of Technology, Gubbi-572216, VTU Belagavi,Karnataka,India:veenasudhi30@gmail.com
[2]Professor,Department of ECE,Channabasaveshwara Institute of Technology, Gubbi- 572216, Tumkur. Karnataka. India:sureshtumkur@yahoo.co.in

ABSTRACT
Quality of life is considered as an important outcome in health research. The information science and technological advances plays a major role in public healthcare. Clinical Decision Support System helps to healthcare professionals to give better diagnostic decisions. Diagnosing Non-Communicable Diseases (NCD) viz., Cardio Vascular Diseases (CVD) and Diabetes Mellitus (DM) required accurate analysis and prediction. To overcome the problems of knowledge based CDSS, machine learning techniques acquire knowledge automatically from the previous patient's clinical data. The techniques used for the diagnosis are depending on a one or more combination of classifiers. The proposed system uses ensemble based methods to give better performance of particular disease prediction. The current ensemble approaches are the enhancement techniques which are based on each stage of outputs.
.
Key words : CDSS, NCD, databases, classifiers.

1. INTRODUCTION

Computational well being informatics is a rising research area which including different sciences like medical, biomedical, nursing, data innovation, software engineering and measurements [3].
CDSS help medical practitioners to diagnose with proper and specific information about patient to enhance health and health care [2]. Using predetermined algorithms and rules.. The rules are configured by the administrator.
CDSS are categorized specifically into major two groups: Non-Knowledge base CDSS and Knowledge base CDSS. These systems consist of Knowledge base, inference rules and a mechanism. Rules are generally IF-THEN statements, Rule, Evidence and Fuzzy systems. Non-Knowledge based systems uses artificial intelligence.

2. RELATED WORKS

Machine learning is an artificial intelligence technique evolved from computational learning theory and pattern recognition. The important techniques of machine learning are Classification and Clustering. Among those few of the significant researches are listed below:
Norul Hidayah Ibrahim et al (2013) proposed a model by combining decision tree and hierarchical clustering for classification of diabetic patients. This model achieved 80.8% accuracy [14].
Kulkarni Rashmi Ravindranath (2015) compared the different algorithm approaches like Naive bayes, Fuzzy logic, SVM, Decision Tree and Neural Network for the diagnosis of heart diseases. The accuracy was achieved by Naive bayes, Fuzzy logic , SVM, Neural Network, and Decision Tree is around 90-95%,78-85%, 85-90%, 80-90% and 94-96% respectively. An implementation of extended sub- tree approaches is used to conquer the problem of ID3 algorithm, which cannot able to continuous data. Among the decision tree is an effective technique for classification [2].
Emarana Kabir Hashi et al (2017) proposed an expert CDSS for the prediction of Diabetes. The system used K-Nearest Neighbour (KNN) and Decision Tree (C4.5) algorithms for the diagnosis. The correctly classified accuracy for C4.5 is 90.43%, KNN is 76.96%. The C4.5 gives better accuracy than KNN [3].
Sid Ahmed Mokeddaem (2017) developed a CDSS to diagnose Cardio uses C5.0, RF algorithm and fuzzy modeling. The RF algorithm is used for order the features, C5.0 are used for crisp value generation. Then fuzzy weighted rules generated based on crisp values. The experiment results show that the accuracy was around 90.50% [1].
The Neural Network (NN) is a subset of machine learning. NN have been extensively applied to nonlinear statistical modeling problems. To develop the inter relationship between the diagnosis and symptoms, Neural Network uses nodes with weighted interconnections.
Mohammad A M et al (2014) designed an automated system for heart disease diagnosis by combining ANFIS with ANN using MATLAB. Based on experimental work, the accuracy of ANFIS (100%) has better than ANN (90.74%) for training data, whereas for testing data ANN (87.04%) performs better than ANFIS (76.93%) [11].
Jung-Gi Yang et al (2014) developed a hybrid prediction model for CVD by using ANFIS and Linear Discriminate Analysis (LDA), which gives 80.2% accuracy. The dataset used for the analysis was the KNHNES V [6].

Genetic algorithm is based on evolutionary process. The genetic system uses an iterative procedure for feature selection, the process of selecting the most relevant inputs for a predictive model. These techniques will remove irrelevant and redundant features which will not effecting the accuracy of the predictive model.

Aishwarya S et al (2014) designed a system for diagnosis of diabetes based on LS-SVM and Genetic algorithm . The system uses 10-fold cross validation for the analysis. The accuracy of the system was 81.33% [10].

3. FINDINGS AND DISCUSSIONS

To build a clinical decision support systems, literature survey describes a various researches based on Machine Learning, Neural Networks and Genetic algorithms. The Cleveland dataset is the most relevant database used for testing and training the system for heart diseases. The Pima Indians dataset is the main database for diabetes. The datasets and selected attributes for classification are listed in the Table I

Table 1: Datasets and their attributes

Dataset	**Problem domain**	**Attributes**
Cleveland	Cardio Vascular Diseases [14 attributes]	Sex ,Age, CP, Chol ,trestbps, fbs, exang, Slope,restecg, thalach, , oldpeak, , ca, Num, thal
Pima Indians dataset	Diabetes [9 attributes]	Age , times_ pregnant, glucose_cone, Diastolic_ bp (mm Hg), Triceps(mm), serum (mu U/ml BMI, Diabetes _func Class Variable
Hungarian dataset Hungarian Institute of cardiology, Budapest	Heart diseases	Same as Cleveland dataset
Bahrain Defense Force Hospital (BDFH)	Heart diseases and Diabetes	Same as Cleveland dataset 23,000 patients

Algorithms play a major role in the performance of the system. Some of the frequently used algorithms used for the construction of the system are listed along with problem domain in Table 2.

Table 2: Classification Algorithms

Application Domain	**Algorithms**
Cardio Vascular diseases	BPNN algorithm
	Fuzzy Logic approach
	Naive bayes classifier
	Decision tree, CART C4.5, C5.0
	Support Vector Machine
	ANN & ANFIS-LDA (Hybrid system)
	MLPNN and ANFIS (Hybrid system)
	Random Forest algorithm
Diabetes	Decision Tree C4.5, C5.0, CART, J48
	KNN
	SVM
	Bayesian network
	Decision Tree& KNN (Hybrid system)
	ANN and FNN (A hybrid NN)
	LS-SVM and GA (Hybrid model)

4. RESULT ANALYSIS

The system performance analysis is calculated using four metrics like Confusion matrix, Classification accuracy, Specificity and Sensitivity [10][4][16].

Confusion matrix is a 2x2 matrix which illustrates the actual and predicted classification as shown in Table 3.

Table 3: Confusion Matrix

		Actual	
		Negative	**Positive**
Predicted	**Negative**	True Negative	False Negative
	Positive	False Positive	True Positive

The different techniques used for the prediction of Non Communicable Diseases Viz., Diabetes and Cardio Vascular diseases with prediction accuracy are shown below in Table 4.

Table 4: Performance analysis

Author	Techniques Used	Accuracy
Gwenolequellec et al [17]	Neural Network Back propagation algorithm	80-90%
Kittipol Wisaeng [17]	Fuzzy logic approach Weight applied on IF-THEN rules	78-85%
Shamsher Bahadhr patil	SVM, Maximum and Optimum margins Gaussian theorem	85-90%
Kulkarni Rashmi Ravindranath et al [2]	Decision tree, C4.5,CART	94-96%
Jung-Gi Yang et al [6]	C4.5	68.60%
Mahammad A M et al [11]	MLPNN & ANFIS Back propagation algorithm	90.74%
Humar Kahramannli et al [9]	ANN and FNN (A hybrid neural network)	86.80%
Sid Ahmed Mokeddem [1]	C5.0 and Random forest algorithm (A fuzzy classification model)	90.50%
Emrana Kabir Hashi et al [3]	Decision Tree (C4.5)	90.73%
	KNN	76.96%
Humar Kahramannli et al [9]	ANN and FNN (A hybrid neural network)	84.24%
Aishwarya S [10]	LS-SVM and GA (Hybrid Model)	81.33%
Ibrahim et al [15]	Decision tree with hierarchical clustering	80.80%
Karthikeyani et al[10]	Decision tree (C4.5) 10-fold cross validation	86.00%

The algorithm with the highest accuracy, sensitivity, and specificity will be selected as the best classification model. The formulas are shown in equations 1,2, and 3 as mentioned below.

Accuracy: It is the percentage of correctly diagnosed to the overall instances in the dataset.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Sensitivity: It is the total number of correctly diagnosed positive instances.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (2)$$

Specificity: It is the total number of correctly classified negative instances

$$\text{Sensitivity} = \frac{TN}{TN+FP} \quad (3)$$

5. CONCLUSION AND FUTURE WORK

Diagnosis of Cardio Vascular Diseases and Diabetes has been considered as one of the major health challenges in medical field. The proposed system uses Ensemble based methods for the better the performance for the prediction of the disease. The current ensemble approaches are the enhancement techniques in the field of machine learning. These techniques are supported each stage of outputs and functionality of the classifiers The system uses different ensemble techniques viz., stacking. bagging and boosting. These techniques are used for the prediction of liver diseases, hepatitis and headache disorders in the medical field.

REFERENCES

[1] Mokeddem, Sidahmed. **"A fuzzy classification model for myocardial infarction risk assessment.**" *Springer*, (2017):.

[2] Ravindranath., Kulkarni., and Rashmi. "**Clinical Decision Support System for Heart diseases using Extended Subtree"**. In: *Int Conf Pervasive Computing (ICPC) IEEE* (2015).

[3] Kabir Hashi, Embrana, Uz Zaman, Md Shahid, and Rkibulhasan, Md. **"An Expert Clinical Decision Support System to Predict Disease using Classification Techniques".** In: *Int conf Electri, Computer &Com Eng (ECCE) IEEE* (2017): 396-400.

[4] Kelth,Brian., YuanLin., and Bayrak, Coskun. **"Comparison of AI Techniques for Prediction of Liver Fibrosis in Hepatitis Patients."** *Springe*r (2014): 38-60

[5] Zhen, Sheng., Cao, Chunxiang., Guanghe Li., et al. **"Incidence prediction of Communication diseases after the Wenchuan Earthquatke using Remote Sensing."** *IEEE* (2012): 927-930.

[6] Gi Yang, Jung., Kim, Jae-Kwon., Kang, Un-Gu.,et al. **"Coronary heart disease Optimization system on adaptive network-based fuzzy inference system and linear discriminant analysis (ANFIS-LDA)."** *Springer* 18 (2014): 1351-1362.

[7] Jing, Si-Yuan., **"A Hybrid Genetic algorithm for feature subset selection in rough set theory."** *Springer* 18 (2014): 1373-1382.

[8] Rathore, Heena., and Saman,Abhayt. **"Mobile network effects on Communicable disease models."** *IEEE* (2012): 126-131.

[9] Allahverdi, Novruz and Kahramanli, Humar. **"Design of a Hybrid System for the Diabetes and Heart diseases."** In: *Expert Systems with Applications* 35 (2008): 82-89.

[10] S, Anto., and S, Aishwarya**. "A Medical Decision support System based on Genetic algorithm and Least Square Support Vector Machine for diabetes Disease Diagnosis."** In: *Int J Engg Sci & Research Tech* (2014): 4042-4046.

[11] A, M, Mohammad., Abushariah., A, M, Assal., et al. **"Automatic heart Disease Diganosis System Based on Artificial Neural Network(ANN) and Adaptive**

Neuro-Fuzzy Inference (ANFIS) approaches." In: *J software Eng & Applications* 7 (2014): 1055-1064.

[12] D, C, Bindushree., and Dr.V. Udayarani. **"A Review on using various DM techniques for Evaluation of performance and analysis of heart disease prediction."** *IEEE* (2017): 686-690.

[13] Aladallal, Ammar., Abdul Aziz, Amina., and Al-moosa. **"Using dataming technique to predict diabetes and heart diseases ."** *IEEE* (2018): 150-154.

[14] K. K. Ghandhi., and N. B. Prajapathi. **"Diabetes prediction Using feature selection and classification." In:** ***Int J Advance Eng & Research Development*** **1(5) (2014): 2348-6406**

[15] A,Mustapha., N, H, Ibrahim., et al. **"A hybrid model of hierarchical clustering and decision tree for rule based classification of diabetic patients."** In: *Int J Eng and Tech (IJET)* 5(5) (2013): 3986-3991.

[16] V, Kartikeyani., I, P, Begum., et al**. "Comparative of datamining classification algorithm(CDMCA) in diabetes disease prediction."** In: *Int J Computer applications* 60(12) (2012): 26-31.

[17] Quellec, Gwenole., Lamard, Mathieu., et al. **"Medical case retrieval from a Committee of Decision tree."** *IEEE* 14(5) (2011): 1227-1235

.