Deep Learning based Model for Decision Support with Case Based Reasoning

Luxmi Verma, Manish Kumar Mathur

Abstract: ***Cardiovascular diseases (CVD) the most common reason for deaths worldwide, are not easily diagnosed in initial stages. Early and accurate detection of CVD is highly required to prevent this leading cause of mortality. In the last few years, with the advancement of technology and increased potential of digital procedures, almost every business sector is adapting the automation and hence generating a large volume of data. Health sector is also affected by this outburst of technology and almost all the hospitals are generating a huge volume of data every day. The need of the hour is how to handle such a huge data and finding the hidden correlations among it so that it can be used by clinical experts in disease diagnosing and helps them in decision-making. This paper presents an intelligent decision support model for detection of coronary artery disease (CAD) with the integration of cuckoo algorithm for feature subset, analysis of various classification techniques to diagnose the disease more accurately and case base reasoning (CBR) for detecting the severity of the disease. The results seems promising and the integrated technique shows the accuracy of MLP is 85.48 %.The model can be used as a promising decision making tool for medical experts for detecting cardio vascular diseases in their early stages.***

Index Terms : Cardiovascular Disease, Multilayer Perceptron, Case Based Reasoning

I. INTRODUCTION

Knowledge discovery from the huge amount of data, which is being generated every day, is one of the biggest challenge in front of the healthcare industry. It is not possible to analyze such a vast amount of data manually, so medical practitioners needs some kind of intelligent decision support system for diagnosing the diseases based on clinical parameters. Conventional statistical decision models are not capable enough to handle large, complicated and nonlinear dependent [1]. They require a different approach towards data management. The integration of statistical methods of modelling with other technologies such as machine learning and case based reasoning can yield more promising results.
Cardiovascular disease are the biggest threat to any country and were once considered to be affecting only the developed countries, but it is now proved that it constitutes a major portion of non-communicable diseases in the developing countries also. Figure 1 shows the shift of disease burden in the next 12 years and figure 2 displays the major causes of mortality by 2030 [2]. The major reason for CVD is the accumulation of plaque in the arteries which narrow down the arteries and restricts the flow of blood [3-5].

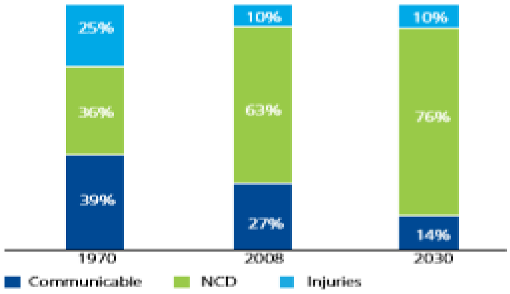

Figure 1: Global burden of diseases

The most promising method for detection of CAD is angiography, but it is expensive, painful and requires a technical expertise. Such complications and limitations leads to the invention of other non-invasive methods. In the last few years machine learning has gained a lot of attention in almost all spheres of life because of its ability to find out complex correlations between data and amazing computational abilities.
In this paper we propose a model that has capacity to accept a user query, evaluate the query parameters as test case and predicts whether the patient is suffering from CAD or not and if he is positive to CAD then the system retrieve matching case from CBR to know the severity of the disease. The proposed model would reduce the detailed analysis of huge data and minimizes the efforts of finding the related patterns repeatedly for the similar types of problems thus reducing the complex knowledge base decision-making process.

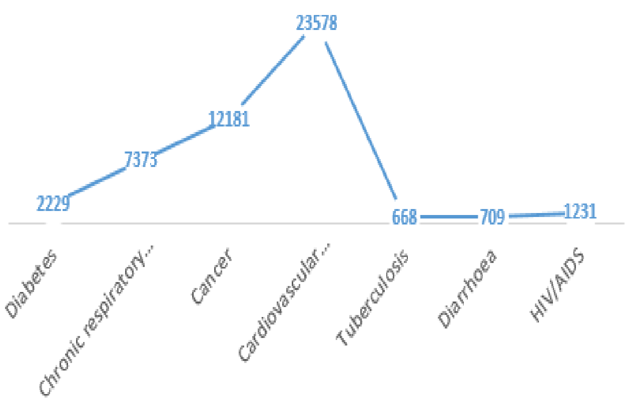

Figure 2. Major causes of deaths by 2030

Many researchers in the past have used many intelligent methods for the diagnosis of medical disease and decision support systems. Yeow et al. (2014) proposed CBR as one of the promising tool for forensics.

Revised Manuscript Received on December 22, 2018.

Luxmi Verma, Department of Computer Science & Engineering, Devbhoomi Institute of Technology, Dehradun, India .
Manish Kumar Mathur, Devbhoomi Institute of Technology, Dehradun, India

They highlighted the limitations of existing forensics methods and a novel use of CBR as a method for analysis of forensic evidences. They have used the methods of information extraction for abstracting the contents from autopsy reports and then find the similar cases by using CBR and Naïve Bayes technique. The proposed system provides reasonably correct results and outshined other traditional methods of forensics [6].

Anggrawan et al. (2016) in their study suggested a framework for the diagnosis of kidney failure. Authors used php and mysql for the implementation of the framework and used matching coefficient to calculate the similarity between the old and new cases. Their model showed the accuracy of 80% that has been tested using spearman rank test [7].

Singh et al. (2016) proposed a model for the diagnosis of Psychiatric Disorder. They have used 750 cases and out of which 500 cases have been used for initial case base creation and rest 250 were used as a testing data. The used k-nearest neighbor method for matching the index value of the case with those that have been stored in the case base. Their model achieved the accuracy of 96% [8].

Babic et al (2017) worked on three different data sets: South African heart disease, Heart disease database and dataset from Z-Alizadeh Sani. They had implemented Neural Networks, Naïve Bayes, SVM and Decision Trees. Further, the author performed descriptive analysis based on association and decision rules. The models proposed by authors are comparable with existing studies and in some cases comparable or better [9]. Pandey et al. (2017) in their study proposed a medical diagnosis system for ECG based diseases. The author provided an integration system of a classification approach J48 and a CBR component. Their study improved the diagnostic accuracy to 96% [10]. Vedayoko et al. (2017) in their work focused on implementation of expert system integrated with CBR for the diagnosing of bowel disease. They used k nearest neighbor algorithm for finding the similar cases in CBR component. They have implemented the model on the data collected from RSUD dr. Soetrasno Rembang Regency for 60 patients, who were having bowel diseases like Gastroentritis, Tipoid Fever, Diarrhea and Colitis [11].

II. DATA DESCRIPTION

Cleveland Heart disease data set is obtained from one of the most popular repository of the University of California at Irvine [12]. It consists of features like Gender, Age, Type of chest pain, Resting electrocardiographic outcome, Serum cholesterol, Maximum heart rate achieved, Fasting blood sugar, blood pressure (Resting) on admission, Exercise induced angina, Number of fluoroscopy colored vessel, ST depression induced by exercise related to rest, Slope of the peak exercise ST Segment, Thal and result of Angiography as a deciding factor for CAD.

III. METHODOLOGY AND LEARNING SCHEMES

Preprocessing of the data was carried out using feature subset selection with Cuckoo search method [13]. Then models (Random Forest, ANN, FURIA and Logistic regression) were created with supervised learning were validated with ten-fold cross validation method. The performance measures that were recorded are misclassification error rate, accuracy, Kappa Statistics and Mean absolute error.

Table1: Description of dataset

Features	Description	Range (Min Max)		Mean	StDev
Age	Age (in yrs)	20	77	55.43	9.03
Sex	0-female, 1-male	0	1	0.68	0.46
Cp	Type of chest pain 1- typical Angina 2- atypical angina 3- non-angina pain 4- asymptomatic	1	4	3.15	0.96
Trestbps	blood pressure (resting) on admission.	90	200	131.69	17.6
Chol	Serum cholesterol mg/d	126	564	246.6	51.77
Fbs	Fasting blood sugar > 120 mg/dl 0 - no 1- yes	0	0-1	0.149	0.356
Restecg	Resting Electrocardiographic outcome 0-normal 1- having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) 2: showing probable or definite left ventricular hypertrophy by Estes' criteria	0	0-2	0.99	0.995
Thalach	Max heart rate achieved	71	202	149.60	22.8
Exang	Exercise induced angina 0 = no 1 = yes	0	1	0.327	0.47
Old peak	ST depression induced by exercise related to rest	0	6.2	1.04	1.161
Slop	Slope of the peak exercise 1-upsloping 2- flat 3- down sloping	1	3	1.601	0.616
Ca	No. of fluoroscopy colored vessels	0	3	0.672	0.937
Thal	3 – normal 6 -fixed defect 7 -reversible defect	3	7	4.73	1.94

Stage 1 : Data preprocessing

Reducing dimensionality-using Correlation based feature subset selection and Cuckoo search technique.

Stage 2 : Construction of Model and validation

(Deep learning, Random forest, FURIA, Bagging by means of10 fold cross validation)

Stage 3 :Performance measure Accuracy, Misclassification error rate, Mean absolute error, Kappa Statistics were used to evaluate.

Stage 4: The positive identified case is then fed into CBR for finding the best possible match for the severity of the disease.

Figure 3. Framework for Model construction

A. The Proposed Model

Case base reasoning (CBR) is a kind of analogous reasoning that is widely used in decision making [14-16]. CBR is an effective technique for handling decision based problems and used in all almost all the spheres of business. It helps businesses to take decision based on the previous experiences encountered and provide the best possible matching scenario. CBR uses simple matching coefficient, nearest neighbor and naïve Bayes techniques to find the closest match [17-18]. It has been successfully applied to various industrial applications like sales operations, quality management, product development. The major reason for the success of CBR is its ability to reason with the situation based on the previous experiences.

CBR follows a systematic cycle, which consists of four R's:

Retrieve: the phase deals with the process of finding the most similar cases.

Reuse: It explains how the retrieved case(s) can be used to solve the problem statement.

Revise: It is the proposed solution for the problem based on the retrieved case.

Retain: The storing of new case after revision in the case base for future use.

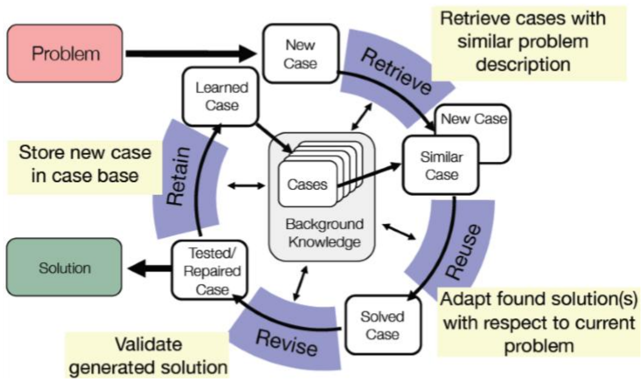

Figure 4: CBr Cycle [19]

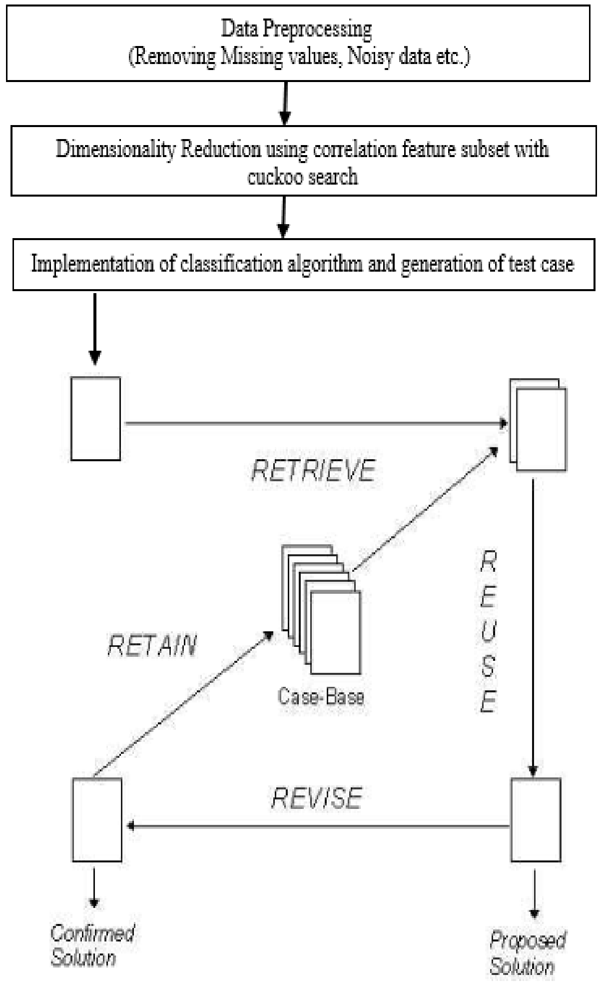

Figure 5. Proposed Model

IV. RESULTS

The models were constructed using Cleveland dataset for Heart Disease gathered from UCI Machine repository, with Deep learning [20], Random Forest, Fuzzy Unordered Rule Induction and bagging Classifiers. The constructed models were assessed for error rate, accuracy, Kappa Stat, Mean Absolute error and RMS (Root Mean Square error). It is found that Deep learning based model with Correlation based feature subset selection with PSO search achieves the maximum accuracy of 85.481% which is highest among other classifiers. Comparative results are shown in (Table 2).

Table 2: Accuracy/error rate/MAE/RMSE

Model	Accuracy	Error	KS	MAE	RMSE
Deep learning	85.48	14.5	0.6267	0.2015	0.3333
Random Forest	81.84	18.15	0.5351	0.2196	0.3414
FURIA	79.53	20.46	0.4856	0.21	0.429
Bagging	82.178	17.82	0.5418	0.2443	0.3505

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table 3. Accuracy/error rate/KS/MAE/RMSE with all the features

Model	Accuracy	Error	KS	MAE	RMSE
Deep learning	83.16	16.83	0.5623	0.2065	0.3357
Random Forest	80.198	18.15	0.5351	0.2196	0.3414
FURIA	78.8779	20.46	0.4856	0.21	0.429
Bagging	80.198	19.80	0.4791	0.2645	0.3667

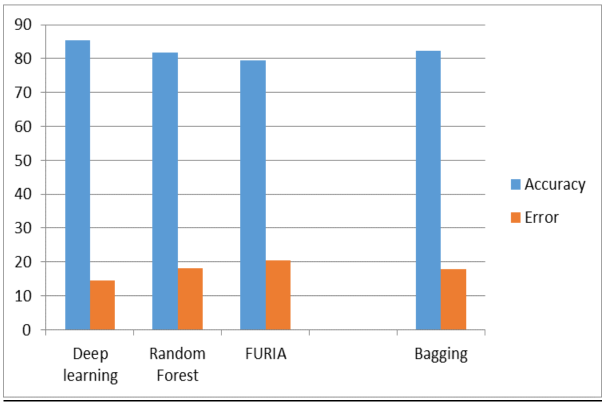

Figure 6: Accuracy and error rate of models with feature selection

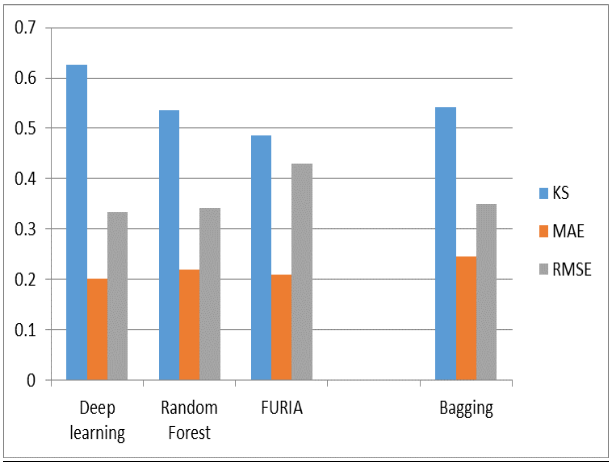

Figure 7: KS/MAE/RMSE of models with feature selection

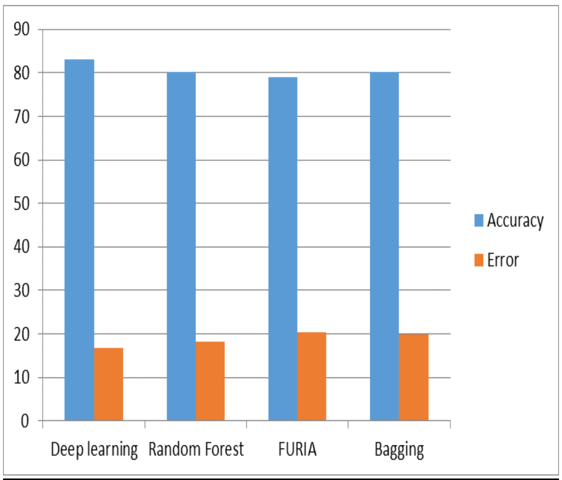

Figure 8: Accuracy and error rate of models with all feature selection

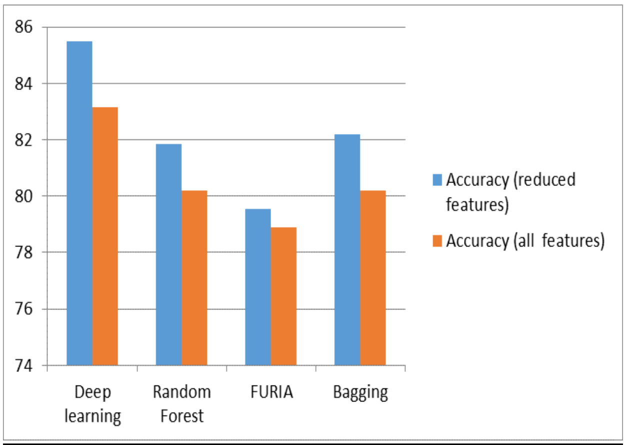

Figure 9. Accuracy of Models by using all the features and with dimensionality reduction method

V. DISCUSSION & CONCLUSION

The study implemented proved that the deep learning method outshined all the other techniques used, for detection of Coronary heart disease with correlation based feature subset selection method with cuckoo search. Dimensionality of the data set is reduced and only few parameters are abstracted to construct the model. The most influential risk factors identified to predict the CAD are CP, Thalach, Exang, oldpeak, slope, CA, Thal. Deep learning based method achieved the highest prediction accuracy of 85.48% with error rate as low as 14.5% and highest the value of Kappa Statistics i.e. 0.6267 and lowest root mean square error i.e. 0.3333. By reducing the dimensionality of the feature space the accuracy of the deep learning method is increased by 2.23 %. Random forest attains the prediction accuracy of 81.84% with error rate of 18.15%. The value of KS for random forest is 0.5351 and root mean square error is 0.314. On the other hand, FURIA showed the lowest prediction accuracy of 79.52% and error rate of 20.46%. Value of Root mean square error is highest in case of FURIA. Bagging achieves the prediction accuracy of 82.178%, error rate of 17.82%. Value of KS for Logistic regression is 0.5418 and value of root mean square error is 0.3505. The results of deep learning model were used as a new case for the CBR and CBR achieved the similarity accuracy of 89% from the previous experiences stored in it. Thus, deep learning based model integrated with CBR component can be used to predict CAD cases more accurately.

REFERENCES

1. Lin, R. H. (2009). An intelligent model for liver disease diagnosis. Artificial Intelligence in Medicine, 47(1), 53-62.
2. International Heart Protection Summit, September (2011) Cardiovascular diseases in India: Challenges and way ahead. India: ASSOCHAM.
3. El-Bialy, R., Salamay, M. A., Karam, O. H., & Khalifa, M. E. (2015). Feature analysis of coronary artery heart disease data sets. Procedia Computer Science, 65, 459-468.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

4. Chung, J. (2017), Association between Carotid Artery Plaque Score and SYNTAX Score in Coronary Artery Disease Patients. General Medicine: Open Access, 5(5).
5. Zhou, H., Wang, X., Zhu, J., Fish, A., Kong, W., & Li, F. et al. (2017). Relation of Carotid Artery Plaque to Coronary Heart Disease and Stroke in Chinese Patients: Does Hyperglycemia Status Matter?. Experimental And Clinical En-docrinology & Diabetes, 126(03), 134-140.
6. Yeow, W. L., Mahmud, R., & Raj, R. G. (2014). An application of case-based reasoning with machine learning for forensic autopsy. Expert Systems with Applications, 41(7), 3497-3505.
7. Anggrawan, A., Hidjah, K., & Jihadil, Q. S. (2016, October). Kidney failure diagnosis based on case-based reasoning (CBR) method and statistical analysis. In Informatics and Computing (ICIC), International Conference on (pp. 298-303). IEEE.
8. Singh, P., Singh, A. P., & Ahmad, S. (2016, October). Case based reasoning model in the diagnosis of psychiatric disorder. In Communication and Electronics Systems (ICCES), International Conference on (pp. 1-6). IEEE.
9. Babič, F., Olejár, J., Vantová, Z., & Paralič, J. (2017, September). Predictive and descriptive analysis for heart disease diagnosis. In Computer Science and Information Systems (FedCSIS), 2017 Federated Conference on (pp. 155-163). IEEE.
10. Pandey, B., & Kundra, D. (2017). Diagnosis of EEG-based diseases using data mining and case-based reasoning. International Journal of Intelligent Systems Design and Computing, 1(1-2), 43-55.
11. Vedayoko, L. G., Sugiharti, E., & Muslim, M. A. (2017). Expert System Diagnosis of Bowel Disease Using Case Based Reasoning with Nearest Neighbor Algorithm. Scientific Journal of Informatics, 4(2), 134-142.
12. https://archive.ics.uci.edu/ml/index.php
13. Yang, X. S., & Deb, S. (2014). Cuckoo search: recent advances and applications. Neural Computing and Applications, 24(1), 169-174.
14. Richter, M. M., & Weber, R. O. (2016). Case-based reasoning. Springer-Verlag Berlin An.
15. Bichindaritz, I., & Marling, C. (2006). Case-based reasoning in the health sciences: What's next?. Artificial intelligence in medicine, 36(2), 127-135.
16. Wang, W. M., Cheung, C. F., Lee, W. B., & Kwok, S. K. (2007). Knowledge-based treatment planning for adolescent early intervention of mental healthcare: a hybrid case-based reasoning approach. Expert Systems, 24(4), 232-251.
17. Xu, W., Xiong, G., Gao, F., & Zhang, X. (1999). Case based reasoning in conflict negotiation in concurrent engineering. Tsinghua Science and Technology, 4(2), 1397-1402.
18. Tsai, C. Y., & Chiu, C. C. (2009, April). Developing a Significant Nearest Neighbor Search Method for Effective Case Retrieval in a CBR System. In Computer Science and Information Technology-Spring Conference, 2009. IACSITSC'09. International Association of (pp. 262-266). IEEE.
19. Bach, K., Sauer, C., Althoff, K. D., & Roth-Berghofer, T. (2014, August). Knowledge modelling with the open source tool myCBR. CEUR Workshop Proceedings.
20. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural networks, 61, 85-117.