An Ensemble Deep Dynamic Algorithm (EDDA) to Predict the Heart Disease

J. Nageswara Rao[1], Dr. R. Satya Prasad[2]

[1]Research Scholar Department of Computer Science and Engineering, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India

[2]Professor, Department of Computer Science and Engineering, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India

ABSTRACT

Nowadays heart disease becomes more complicated to every human being. Machine Learning and Deep Learning plays the major role in processing the automatic systems. Prediction of heart disease is most difficult task because many algorithms perform limited operations. The aim of the paper is to increase the accuracy and prediction values. Various heart disease datasets are available for the research. Deep Learning (DL) algorithms play the major role in prediction of heart disease. Prediction can be done in the early stages to reduce the risk of death for the humans. In this paper, An Ensemble Deep Dynamic Algorithm (EDDA) is introduced to increase the accuracy of prediction values. The EDDA follows the some steps to process the prediction of heart disease. The steps are as follows: Linear Regression and Deep Boltzmann Machine (DBM) is applied on the selected dataset. Performance is calculated in terms of sensitivity, specificity and accuracy are shown with the comparative results.

I. INTRODUCTION

In the present world, there is no death of records regarding symptoms of patients that are suffering with heart disease. In India, as per the Indian Heart Association, less than 50 years of age people 50% and under 40 years 25% people are affecting with heart disease for Indians. Many of the urban people are 3 times more affecting with heart attacks as rural population. Heart disease prediction can be observed by various symptoms such as breathing problem, age factor, blood pressure etc causes the heart attacks or heart disease.

Machine learning is an emerging subfield of artificial intelligence (AI). The purpose of ML is to design a system that allows learning and making predictions based on previous data. It uses training data sets to train machine learning algorithms to create models. The model uses new input data to predict heart disease. Using machine learning, it can detect hidden patterns in the input data set to build a model. It can make accurate predictions on new data sets. Clear the

data set and fill in missing values. The model uses new input data to predict heart disease and then performs an accuracy test.

In this paper, the Ensemble Deep Dynamic Algorithm (EDDA) is developed to improve the accuracy and other parameters.
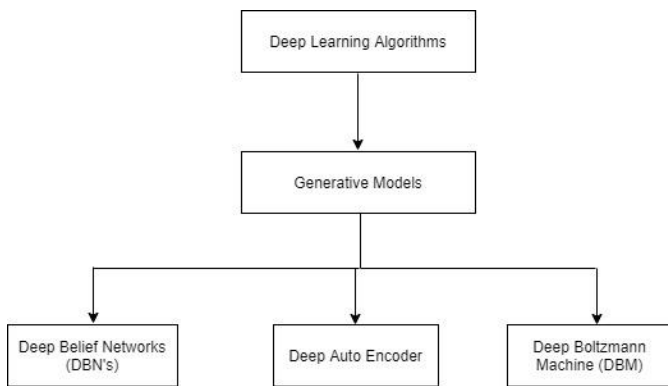

Figure 1 : Boltzmann Machine Algorithms

II. LITERATURE SURVEY

The author SenthilKumar et.al [1], Machine studying (ML) has been verified to be compelling in assisting with deciding on alternatives and forecasts from the sizable quantity of statistics brought via way of the scientific offerings industry. ML processes are being applied in ongoing improvements in diverse zones of the Internet of Things (IoT). Different examinations provide only a quick look at foreseeing coronary contamination with ML techniques. Creators proposed a singular approach that objectives locating important highlights via way of making use of AI processes bringing approximately enhancing the exactness within side the expectation of cardiovascular infection. The forecast version is offered with diverse blends of highlights and some acknowledged characterization methods. We produce an upgraded execution stage with a precision stage of 88.7% thru the forecast version for coronary contamination with the go breed arbitrary backwoods with an linear model (HRFLM).

Anjan Nikhil Repaka, Sai Deepak Ravikanti and Ramya G Franklin[2] applied Naive Bayesian approach to devise and actualize coronary contamination expectation. To accomplish this SHDP (Smart Heart Disease Prediction) is fabricated thru Navies Bayesian to foresee chance elements regarding coronary contamination. The rapid development of innovation has precipitated noteworthy ascent in transportable well being innovation that being one of the net application. The important statistics is accumulated in a normalized structure. For waiting for the percentages of coronary contamination in a patient, the accompanying houses are being added from the scientific profiles, those include: age, BP, cholesterol, sex, glucose and so on The collected traits is going approximately as contribution for the Navies Bayesian association for foreseeing coronary contamination. The dataset used is a part into segments, 80% dataset is used for making ready and relaxation 20% is used for testing. The proposed technique carries following stages: dataset assortment, patron enlistment and login, characterization by Navies Bayesian, expectation and steady statistics pass via way of utilising Advanced Encryption Standard. From that factor end result is created. The exploration explains and offers exceptional statistics deliberation techniques via way of utilising statistics mining strategies which might be acquired for coronary contamination expectation. The yield uncovers that the installation demonstrative framework viably enables with waiting for chance elements regarding coronary heart diseases.

As indicated via way of Ed-Daoudy[3], early discovery of coronary heart infections and ceaseless watching can decrease the dying rate. The dramatic improvement of statistics from diverse sources, for example, wearable sensor devices applied in Internet of Things well being checking, net primarily based totally framework and others were developing a tremendous degree of statistics on a continual premise. The combo of streaming sizable statistics exam and AI is an development innovation which could have a

large impact in hospital treatment area mainly early reputation of coronary contamination. This innovation may be all of the greater notable and greater affordable. A steady coronary contamination forecast framework has been proposed depending on apache Spark. The framework contains of number one sub parts, particularly streaming making ready and statistics stockpiling and perception. The most important makes use of Spark MLlib with Spark streaming and applies order version on statistics events to foresee coronary contamination. The seconds makes use of Apache Cassandra for placing away the sizable extent of produced statistics.

Amin Ul Haq et.al[4]have consider Hybrid Intelligent System Framework for the Prediction of Heart Diseases. The creators attested that noninvasive-primarily based totally strategies, for example, AI are stable and proficient. A Machine-studying-primarily based totally end framework for coronary contamination forecast via way of utilising coronary contamination dataset turned into created utilising seven widely known AI calculations, 3 aspect dedication calculations, the go-approval technique, and 7 classifiers execution evaluation measurements, for example, association exactness, particularity, affectability, Matthews' courting coefficient, and execution time. The proposed framework can certainly apprehend and institution people with coronary contamination from sound people. Additionally, receiver optimistic curves and area under the curves for each classifier was computed. The authors permitted execution of the proposed framework on complete highlights and on a reduced association of highlights. The highlights lower impacts classifiers execution concerning exactness and execution season of classifiers.

Md. Shahriare Satu et.al [5] gift that Heart Disease is one of the most important ailments that reasons brilliant lack of lives anywhere at the world. Some unordinary methods to address find out important factors of coronary heart diseases were taken into

consideration via way of the creators. They have applied coronary contamination statistics (Cleveland and Hungarian) and the 2 of them are partitioned into 33%, 65% and 100% statistics. Estimations of diverse scope of man or woman credit in those statistics are resolved to find out large additives of this sickness. At that factor, special semi administered studying calculations, for example, Collective Wrapper, Filtered Collective however Another Semi Supervised Idea are applied to research coronary contamination statistics. Measurements of those classifiers like precision, f-degree and location beneathneath ROC were decided to legitimize singular classifiers and suggest the nice semi controlled studying calculation. This calculation is investigated large and superfluous variables of coronary contamination via way of doing away with credit continuously successively and noticing the effects of association. Test effects on actual statistics showcase the adequacy and productiveness of the exam.

III. Dataset Description

The data set used is the "heart disease data set", which is a combination of 4 different databases, but only the UCI Cleveland data set is used. The database has 76 attributes in total, but all published experiments only use a subset of 14 features to represent [9]. Therefore, we used the processed UCI Cleveland dataset provided on the Kaggle website for analysis. Table 1 below lists a complete description of the 14 attributes used in the proposed work.

Sl.No.	Attribute Description	Distinct Values of Attribute
1.	Age- represent the age of a person	Multiple values between 29 & 71
2.	Sex- describe the gender of person (0- Female, 1-Male)	0,1
3.	CP- represents the severity of chest pain patient is suffering.	0,1,2,3
4.	RestBP-It represents the patients BP.	Multiple values between 94& 200
5.	Chol-It shows the cholesterol level of the patient.	Multiple values between 126 & 564
6.	FBS-It represents the fasting blood sugar in the patient.	0,1
7.	Resting ECG-It shows the result of ECG	0,1,2
8.	Heartbeat- shows the max heart beat of patient	Multiple values from 71 to 202
9.	Exang- used to identify if there is an exercise induced angina. If yes=1 or else no=0	0,1
10.	Old Peak- describes patients depression level.	Multiple values between 0 to 6.2.
11.	Slope- describes patient condition during peak exercise. It is divided into three segments(Unsloping, Flat, Down sloping)	1,2,3.
12.	CA- Result of fluoroscopy.	0,1,2,3
13.	Thal- test required for patient suffering from pain in chest or difficulty in breathing. There are 4 kinds of values which represent Thallium test.	0,1,2,3
14.	Target-It is the final column of the dataset. It is class or label Colum. It represents	0,1

Table.1 Selected Cleveland Heart Disease Data Set

IV. LINEAR REGRESSION

Linear Regression (LR) is supervised machine learning algorithm that predicts the output is continuous and has a constant slope. Linear regression is the next step up after correlation. It is used when we want to predict the value of a variable based on the value of another variable. The variable we want to predict is called the dependent variable.

1. Start (analyzing the dataset)
2. Read Number of Data (n) (total number of records and attributes)
3. For i=1 to n:
    Read Ai and Bi (//affected or not affected)
  Next i
4. Initialize:
    sumA = 0
    sumA2 = 0
    sumB = 0
    sumAB = 0
5. The SUM is to be calculated
  For i=1 to n:
    sumA = sumA + Ai
    sumA2 = sumA2 + Ai * Ai
    sumB = sumB + Bi
    sumAB = sumAB + Ai * Bi
  Next i

6. Calculate Required Constant a and b of y = a + bx:
  b = (n * sumAB - sumA * sumB)/(n*sumA2 - sumA * sumA)
  a = (sumB - b*sumA)/n
7. Display value of a and b
8. Stop

4.1. Normalization

Normalization is a technique often applied as part of data preparation for machine learning. The goal of normalization is to change the values of numeric columns in the dataset to use a common scale, without distorting differences in the ranges of values or losing information.

4.2. Deep Boltzmann Machine Algorithm

DBM is multilayer structure with undirected graph. This is entirely different from restricted Boltzmann machine (RBM), The DBM consists of number of multiple layers and RBM consists of only one layer. From the observation we can say that DBM is advanced version of RBM. The complex datasets like heart disease prediction can be done by DBM accurately. DBM is the three layer generative model. This is bidirectional connections in the last layers.

$$E(v,h) = -\sum_i^i v_i b_i - \sum_{n=1}^{N}\sum_k h_{n,k} b_{n,k} - \sum_{i,k} v_i w_{ik} h_k - \sum_{n=1}^{N-1}\sum_{k,l} h_{n,k} w_{n,k,l} h_{n+1,l}$$

For a DBM with 'N' hidden layers.

In this paper, initialize the two layers Boltzmann machine within –layer connections. The final state is $\{v, h^1, h^2\}$ is defined as:

$$E(v, h^1, h^2; \theta) = -v^T W^1 h^{1T} W^2 h^2$$

Where $\theta = \{W^1, W^2\}$ are the model parameters, these represents the visible-to-hidden and hidden-to-hidden. The probability that the model gives for the disease prediction is in visible vector v is:

$$p(v;\theta) = \frac{1}{Z(\theta)} \sum_{h^1 h^2} \exp\left(E(v, h^1, h^2; \theta)\right)$$
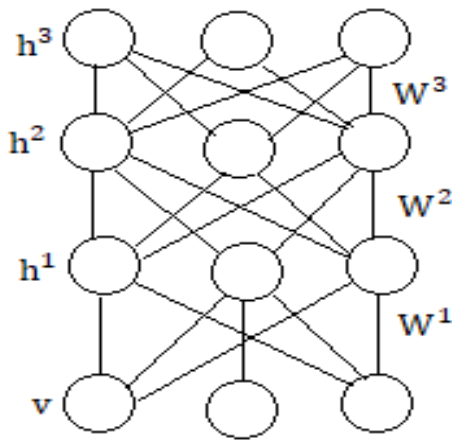
Deep Boltzmann Machine

Figure 3: hidden layers for DBM

Performance Evolution

The performance measures namely False Positive Rate (FPR), False Negative Rate (FNR), Sensitivity, Specificity and Accuracy, the performance of the system are estimated. The basic count values such as True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) are used by these measures.

FPR

The percentage of predicted values was classified to normal and abnormal data, but in fact it did not.

$$FPR = \frac{FP}{FP + TN}$$

FNR

The percentage of predicted values was classified to normal and abnormal data, but in fact it did.

$$FNR = \frac{FN}{FN + TN}$$

Sensitivity

The positives are correctly identified to calculate the sensitivity. This is used to test to identify negative results.

$$Sesitivity = \frac{No.\,of\,TP}{No.\,of\,TP + No.\,of\,TN}$$

Specificity

The negatives are correctly identified to calculate the specificity. This is used to test to identify negative results.

$$Specificity = \frac{No.\,of\,TN}{No.\,of\,TN + No.\,of\,FP}$$

Accuracy: This will calculate the overall accuracy of the abnormal and normal predicted data is calculated by.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

V. Results and Discussion

The implementation is done with Java Programming Language. The NETBEANS 8.0.2 is IDE used to show the results and hardware is 4 GB Ram and 1 TB hard disk. The comparative results are shown in table 1. The parameters based performance is shown in table 1.

Data Mining Techniques	Accuracy	Sensitivity	Specificity
Naive Bayes	91.42%	86.43%	79.76%
Random Forest	89.56%	89.76%	84.54%
Decision Tree	92.32	87.12	86.67
ENDDP	97.98%	97.45%	98.54%
EDDA	98.12	98.45	99.12

Table : 2 comparative result

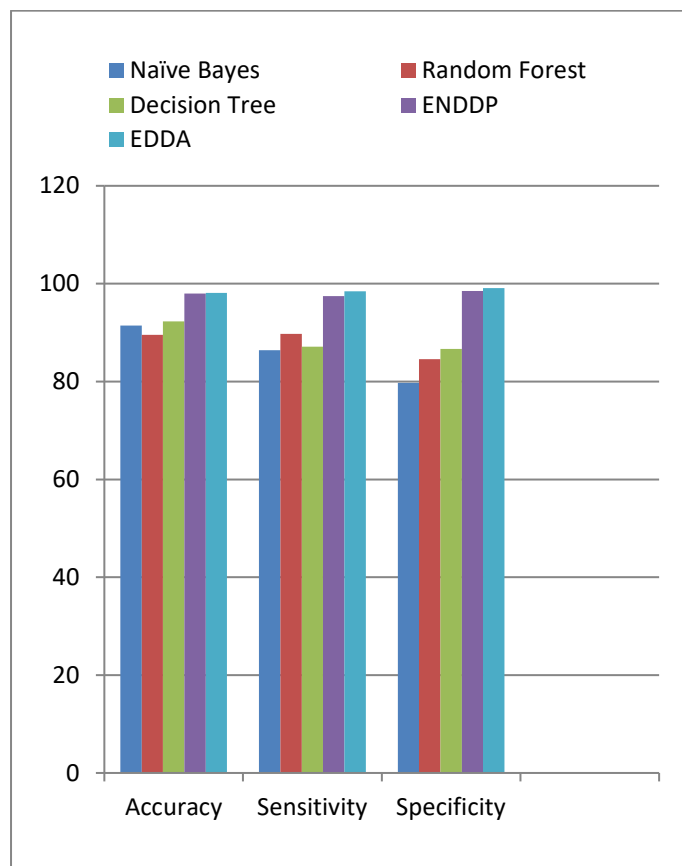This shows the accuracy of the result based on the data mining techniques.

Figure: 4 the performance of the Ensemble Deep Dynamic Algorithm (EDDA)

VI. Conclusion

In this paper, the Ensemble Deep Dynamic Algorithm (EDDA) is implemented by selecting the heart disease dataset which is available in UCI repository. The proposed methodology follows the steps to overcome various issues such as increasing the accuracy and other parameters. The linear regression is model used to train the dataset for improving the performance of the proposed DBM algorithm. DBM is the multilayer hidden based algorithm which is fit for prediction of heart disease.

VII. REFERENCES

[1]. Mohan, Senthilkumar & Thirumalai, Chandra Segar & Srivastava, Gautam. (2019). Effective Heart Disease Prediction using Hybrid Machine Learning Techniques. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2923707.

[2]. A. N. Repaka, S. D. Ravikanti and R. G. Franklin, "Design And Implementing Heart Disease Prediction Using Naives Bayesian," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 292-297.doi: 10.1109/ICOEI.2019.8862604.

[3]. A. Ed-Daoudy and K. Maalmi, "Real-time machine learning for early detection of heart disease using big data approach," 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), Fez, Morocco, 2019, pp. 1-5. doi: 10.1109/WITS.2019.8723839.

[4]. Amin Ul Haq, J. P.Li ,M.H.Memon,Shah Nazir and Ruinan Sun," A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms", Wearable Technology and Mobile Applications for Healthcare, Volume 2018 |Article ID 3860146 | 21 pages | https://doi.org/10.1155/2018/3860146.

[5]. M. S. Satu, F. Tasnim, T. Akter and S. Halder, "Exploring Significant Heart Disease Factors based on Semi Supervised Learning Algorithms," 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), Rajshahi, 2018, pp. 1- 4.doi: 10.1109/IC4ME2.2018.8465642.

[6]. Pahwa K, Kumar R. Prediction of heart disease using hybrid technique for selecting features. In: 2017 4th IEEE Uttar Pradesh section international conference on electrical, computer and electronics (UPCON). IEEE. p. 500–504.

[7]. Pouriyeh S, Vahid S, Sannino G, De Pietro G, Arabnia H, Gutierrez J. A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease. In: 2017 IEEE symposium on computers and communications (ISCC). IEEE. p. 204–207.

[8]. Chauhan R, Bajaj P, Choudhary K, Gigras Y. Framework to predict health diseases using attribute selection mechanism. In: 2015 2nd international conference on computing for sustainable global development (INDIACom). IEEE. p. 1880–84.

[9]. Bouali H, Akaichi J. Comparative study of different classification techniques: heart disease use case. In: 2014 13th international conference on machine learning and applications. IEEE. p. 482–86.

[10]. Xu S, Zhang Z, Wang D, Hu J, Duan X, Zhu T. Cardiovascular risk prediction method based on CFS subset evaluation and random forest classification framework. In: 2017 IEEE 2nd international conference on big data analysis (ICBDA). IEEE. p. 228–32.

[11]. Otoom AF, Abdallah EE, Kilani Y, Kefaye A, Ashour M. Effective diagnosis and monitoring of heart disease. Int J Softw Eng Appl. 2015;9(1):143–56.

[12]. Vembandasamy K, Sasipriya R, Deepa E. Heart diseases detection using Naive Bayes algorithm. Int J Innov Sci Eng Technol. 2015;2(9):441–4.

[13]. Chaurasia V, Pal S. Data mining approach to detect heart diseases. Int J Adv Comput Sci Inf Technol (IJACSIT). 2014;2:56–66.

[14]. Parthiban G, Srivatsa SK. Applying machine learning methods in diagnosing heart disease for diabetic patients. Int J Appl Inf Syst (IJAIS). 2012;3(7):25–30.

[15]. Deepika K, Seema S. Predictive analytics to prevent and control chronic diseases. In: 2016 2nd international conference on applied and theoretical computing and communication technology (iCATccT). IEEE. p. 381–86.