

A Hybrid Technique for Assessment of Heart Sickness Forecast

Navdeep Singh

Research Scholar, Department of Computer Science & Engineering
Shaheed Bhagat Singh State Technical Campus,
Ferozpur (Punjab), India

Sonika Jindal

Department of Computer Science and Engineering
Shaheed Bhagat Singh State Technical Campus
Ferozpur, India

ABSTRACT

The numbers of heart patients with numerous diseases are rising in India each year. The heart issues may be controlled in the initial stages once the on time detection. The on time detection of the heart diseases may be foreseen using the routine health check mechanism for the general public. It becomes a tedious task to investigate the information of heart data of thousands of thousands of patients each week or month. The prediction method needs the expert care professional persons that are terribly less in number. Additionally they kept busy with different route activities like patient checkups, operations, etc. within the many care centers across the country. The matter of in-time prediction may be resolved by using the correct heart prediction formula. The data is often obtainable within the heavier amounts, which may be optimized using the assorted optimization algorithms for the quick and correct process of patient's data. Information has been divided into training and Testing sets to get comparatively higher prediction accuracy. During this paper, the centre illness prediction formula has been designed exploitation the combination of genetic & Naïve Bayes formula for the aim of the data optimization and therefore the result generation. The results from the formula experimentation are obtained exploitation the parameters like accuracy, precision and recall. The results have even the performance improvement of optimized cardiovascular disease prediction solution by using the genetic formula.

General Terms

Machine Learning, Heart Disease, Prediction Analysis, Python, Fuzzy Rules, Feature Selection

Keywords

Classification, GA, Fuzzy Logic, Naïve Bayes

1. INTRODUCTION

In the software system system-engineering field of counterfeit consciousness, a hereditary calculation (GA) is an inquiry heuristic that emulates the procedure of common development. This heuristic is routinely used to turn out helpful answers for improvement and hunt issues. Hereditary calculations slot in with the larger class of transformative calculations (EA), that end up answers for improvement issues utilizing methods roused by characteristic advancement, [8][12] as an example, legacy, transformation, determination, and hybrid. Darwin "s Theory regarding Evolution "Survival of Fittest" enlivens GA. GAs is versatile heuristic inquiry taking into thought the transformative thoughts of common determination and hereditary qualities. GAs is clever misuse of irregular hunt used as a locality of improvement issue. GAs, albeit irregular, misuse authentic data to direct the hunt into the realm of upper execution among the inquiry house.

Hereditary Algorithms (GAs) are versatile heuristic hunt calculation taking into thought the biological process ideas of common various and hereditary qualities. As being what is indicated they speak to a shrewd misuse of an irregular inquiry wont to tackle improvement issues. Albeit irregular, GAs are in no approach, kind or sort capricious, rather they misuse authentic data to direct the inquiry into the realm of upper execution among the pursuit house. The basic methods of the GAs are alleged to recreate forms in common frameworks essential for advancement, notably those take once the standards first set around Charles Robert Darwin of "survival of the fittest." Since in nature, competition among people for brief assets brings regarding the fittest people overwhelming over the weaker ones.

The theorem Classification speaks to a managed learning system and put together a factual technique for grouping. Settle for a hidden probabilistic model and it permits us to catch instability regarding the model in a} very principled manner by deciding probabilities of the results. It'll tackle symptomatic and discerning issues. This Classification is called once Thomas Bayes (1702-1761); World Health Organization planned the scientist Theorem. Theorem characterization offers viable learning calculations and earlier information and watched information are typically joined. Theorem Classification offers a helpful purpose of browse to understanding and assessing varied learning calculations. It figures specific probabilities for speculation and it's powerful to clamor in data information. A Naive Thomas Bayes classifier may well be a basic probabilistic classifier in light-weight of applying Bayes' hypothesis (from theorem insights) with solid (innocent) freedom suspicions. Lots of enlightening term for the elemental chance model would be "autonomous highlight model". [11]

2. LITERATURE REVIEW

Because of a decent accessibility of enormous live of {data of information} and a necessity to change over this accessible large measure of data to valuable data requires the utilization of data mining strategies. Data processing and KDD have gotten to be distinguished as currently. The infamy of data mining and KDD shouldn't be an feeling since the extent of the knowledge accumulations that are accessible are really huge to be analyzed physically and even the techniques for programmed information investigation visible of ancient measurements and machine adapting usually confront issues once transforming huge, dynamic knowledge accumulations comprising of difficult articles. Han and Kamber (2006) described the building squares of today's data processing ways in which come back to Sixties where data and data innovation has been developing methodically from primitive document preparing frameworks to advanced and effective data frameworks. Since the Seventies data frameworks has advanced from right time progressive and system data frameworks to the advancement of social data frameworks

(where knowledge are place away in social table structures), data displaying devices, and categorization and going to strategies. Moreover, clients increased advantageous and convertible knowledge access through inquiry dialects, consumer interfaces, upgraded question transforming, and exchange administration. Heterogeneous data frameworks and Internet-based worldwide data frameworks, for example, the world Wide web (WWW) have likewise developed and assume a key half at intervals the data business (Han and Kamber, 2006). The conversion of orders like AI, knowledge Retrieval, measurements, and information frameworks to boot to the accessibility of quick microcomputers displayed a universe of conceivable outcomes fifteen for recovering and examining data. Amid currently new programming dialects were created and new calculation strategies were created (Dunham, 2003). Naïve mathematician and Neural Network was projected by Palaniappan and Awang, (2008), they utilized a CRISP-DM system to manufacture the mining models on a dataset got from the Cleveland disorder data. The outcomes described the exceptional quality of each of the strategies in appreciating the targets of the predefined mining goals. IHDPS was acceptable noting queries that the customary choice emotionally substantiating networks weren't ready to. It galvanized the inspiration of key learning, e.g. designs, connections at intervals the thick of restorative variables associated with coronary illness. The issue of recognizing compelled affiliation rules for coronary illness forecast was thought of via Carlos the Jackal, (2004). The surveyed dataset fenced restorative records of individuals having coronary illness with traits for danger components, heart movement estimations and passageway narrowing. Three imperatives were at home with abatement the number of examples. Initial one desires the credits to point up to speed stand outside of the rule. The second isolates qualities into uninteresting gatherings Kangwanariyakul et al. (2010) introduced a discerning model for the ischemic disorder (IHD); they connected Back-spread neural system the theorem neural system the probabilistic neural system and conjointly the bolster vector machine (SVM) to create characterization models for recognizing IHD patients on an data no inheriting from estimations of heart engaging field at 36 areas (6 × six grids) over the middle. The top result demonstrates that BPNN and BNN gave the foremost noteworthy characterization truth of 78.43 %, whereas RBF portion SVM gave the littlest quantity order accuracy of 60.78 %. BNN displayed the foremost effective affect-ability of 96.55 makes an attempt to RBF portion SVM showed the foremost-stripped affect-ability of 41.38 %. Every polynomial piece SVM and RBF portion SVM exhibited very cheap and most extreme specificity of 45.45 makes an attempt to eighty-six.36 %, separately.

3. METHODOLOGY

For the implementation of the project we used different algorithms i.e. Naïve Bayes and genetic algorithm. The platform used is Python 3.6.

3.1 Data set

We take the data set in this paper with 303 records and 14 attributes collected from the online dataset repository of archive.ics.edu/ml/datasets. The dataset parameters are listed in Table 1

Table 1: The dataset attributes

Sr. No.	Attributes Name	Description
1	(age)	Patient Age
2	(sex)	Male/Female
3	(cp)	Chest pain type
4	(trestbps)	Resting blood pressure (in mm Hg on admission to the hospital)
5	(chol)	Serum cholesterol (mg/dl)
6	(fbs)	Fasting blood sugar
7	(restecg)	Resting ECG results
8	(thalach)	Maximum heart rate achieved
9	(exang)	Exercise induced angina
10	(oldpeak)	ST depression included by exercise relative to rest
11	(slope)	Slope of the peak exercise ST segment
12	(ca)	Number of major vessels (0-3) colored by fluoroscopy
13	(thal)	3 = normal, 6 = fixed defect, 7 = reversible defect.
14	(num)	Angiographic disease status (Diagnosis of heart disease)

Algorithm 1: GA for Feature Selection

STEP 1: Input dataset

STEP 2: Authenticate the data in respect to make in the appropriate format

STEP 3: Algorithm discovers the best solution for given values and returns the personal best and global best

$$m_g(i=1) = F_g(i) m_g \left[1 - pc \frac{t_g}{t-1} \right] (1 - P_m)^g \quad (1)$$

STEP 4: Compute the pBest and gBest on each module of the dataset for every block

STEP 5: Apply fitness function using Z statistics $Z=S(X)$ -Minimum support/SQRT (min sup*(1-minsup))/N Where S(X) is provision of pattern and min sup is user well-defined level

$$f(x) = \sum_{n=1}^{i=1} [100(x_i + 1 - x_i^2)^2 + (x_i - 1)^2] \quad (2)$$

STEP 6: Cut the features based on Z statistics. After evaluation the rules having maximum fitness are filtered.

3.2 FUZZY RULES

These are the rules that categorize a set of values into user defined output following some Inference Mechanisms. The Inference mechanisms follow such a way, which helps in laying and processing the Fuzzy Rules. While designing the rules, every parameter value must be mapped to an output (should not be an empty set) in order to generate a useful output from a set of values given in the Input.

3.2.1 Inference Mechanisms

Inference mechanism uses the principle of fuzzy logic for synchronizing an input to output using fuzzy logic. It uses membership functions, logical operations and if-then rules.

Madani method works like below:

Consider:

P1: if m is S1 and n is T1 then q is U1

P2: if m is S2 and n is T2 then q is U2

Result: q is U, where m equals m0 and n equals n0.

Sugeno method works like below:

Consider:

P1: if m is S1 and n is T1 then q is q1 = a1m1+b1n1

P2: if m is S2 and n is T2 then q is q2 = a2m2+b2n2

Result: q0, where m equals m0 and n equals n0.

Algorithm2: Cardiovascular Disease Classification Using Naive Bayes

Compute identification="yes", identification="no" P_{yes}, P_{no} from training input.

$$p(C_k | x_1, \dots, x_n) \quad (1)$$

For Every Input Record

Determine Category of Attribute

$$p(C_k | x) = p(C_k) p(x | C_k) / p(x) \quad (2)$$

Compute Probabilities of Identification="Yes", Identification="No"

$$p(C_k, x_1, \dots, x_n) \quad (3)$$

For Each Feature

Compute R_{yes} = R_{yes} * P(Attr, Yes), R_{no} = R_{no} * P(Attr, No);

$$p(C_k | x_1, \dots, x_n) = \frac{1}{z} p(C_k) \prod_{i=1}^n p(x_i | C_k) \quad (4)$$

If (R_{yes} > R_{no}) Then Identify="Yes";

Else Identify = "No";

4. RESULTS AND DISCUSSION

As the goal of this study is to watch heart disease using processing techniques a classification technique was adopted to develop a prognostic model. The models were designed with 2 wholly completely different supervised machine learning algorithms i.e. genetic rule and Naive Bayes. Parameters results have given below. The results are obtained within the form of the varied performance parameters as precision, recall and accuracy percentage shown in Table 2.

Recall = True Positive Rate = Sensitivity

Precision = Positive Predicted Values

True Negative Rate = Specificity = (TN/TN + FP)*100

Specificity = 50%

False Positive Rate = 100 – specificity = 50%

Sensitivity = (TP / TP + FN)*100 = 100
False Negative Rate = 100 – Sensitivity = 100 – 100 = 0%

Table 2: Parameters Results of Proposed Model

Parameter	Accuracy	Precision	Recall
value	97.14%	98%	97.14%

Percentage error: When we calculate results that are aiming for known values, the **percent error** formula is useful tool for determining the precision of your calculations. The formula is given by: The Approximate value is your calculated value, and the exact value is your known value.

((Approximate Value – Exact Value) / Exact Value) * 100
Below is the comparison of percentage error. No of samples have been taken same as in the base paper to avoid any confusion. Values are obtained for each stage separately. Below Table 3 shows that proposed system has less error rate on each stage. So it can be considered more accurate model than previous one.

Table 3: Error Rate at each stage

	No Samples	Of Proposed	Existing
Training	193	17.82	18.65
Validation	59	15.1	15.25
Testing	45	8.56	8.88

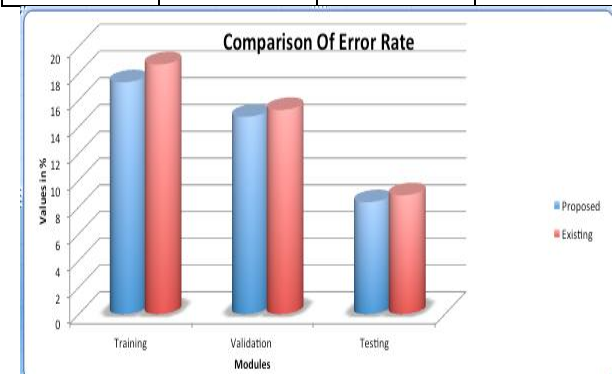


Fig 1: Comparison of Error Rate

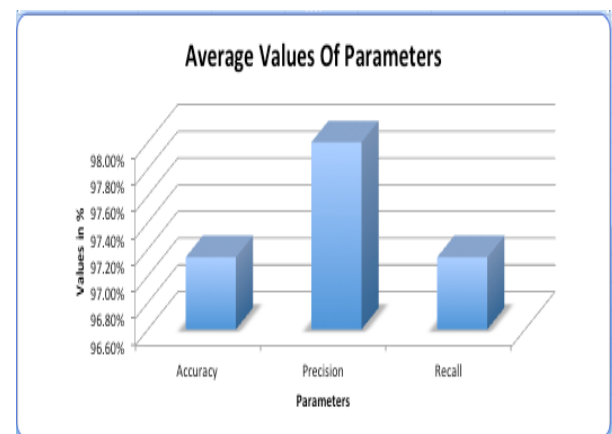


Fig 2: Graph of Proposed Model Results

Obtained Results have been compared with 8 existing systems. Comparison has been shown in Table 3. It has been proved that GA_ Fuzzy_Naive has performed better and

returned higher accuracy than other models. Figure 2 is the graphical representation for Table 4.

Table 4: Comparison of Accuracy with Existing Model

Sr. No.	Model	Accuracy
1	Weighted Fuzzy Rules	57.85%
2	Logistic Regression	77%
3	Naïve Bayes	81.48%
4	Attribute Weighted Artificial Immune System	82.59%
5	Artificial Immune System	84.5%
6	Modified Artificial Immune System	87.43%
7	Neural Networks ensemble	89.01%
8	ANN_Fuzzy_HP	91.10%
9	GA_Fuzzy_Naive (Proposed)	97.14%

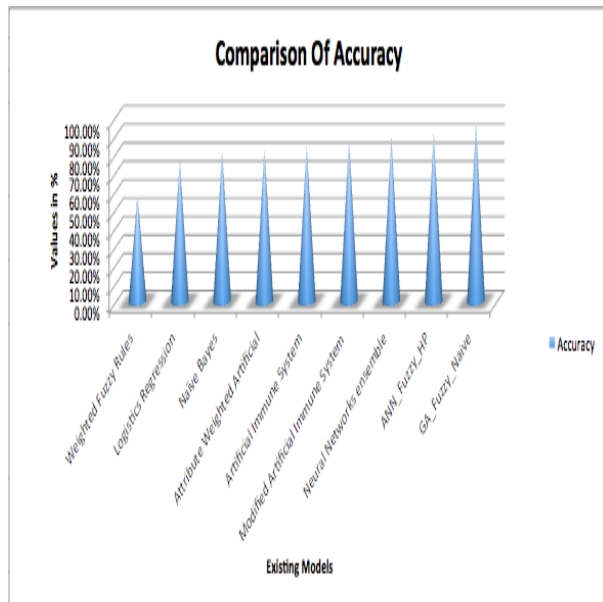


Fig 3: Comparison of Accuracy with Existing Models

5. CONCLUSION & FUTURE SCOPE

The general objective of our work is to forecast a lot of precisely the presence of heart disorder. During this paper, two additional input attributes obesity and smoking are wont to get a lot of correct results. Eight data processing classification techniques were applied. From results it's been seen that GA_Fuzzy_Naive provides correct results as compare to all existing models. This technique may be any dilated. It will use a lot of variety of input attributes listed higher than in table 3 and a pair of. Different data processing techniques may be used for postulation e.g. bunching, statistic, and Connotation rules. The text mining may be wont to mine immense quantity of unstructured information on the market in care trade info. As future work, we are going to method the Categorization and Reversion Trees algorithmic program with different new algorithmic program to refine result a lot of precisely All models could evaluate complex data, each with its own strength with respect to ease of model

interpretation, access to detailed information and accuracy. The relationship between attributes produced by GNBM is more difficult to understand. Proposed Model could be further enhanced and expanded. For example, it can incorporate other medical attributes besides the 15. It can also incorporate other data mining techniques eg Time Series, Clustering and Association Rules.

6. REFERENCES

- [1] Mohammad Taha Khan, Dr.ShamimulQamar and Laurent F. Massin, A Prototype of Cancer/Heart Disease Prediction Model Using Data Mining, International Journal of Applied Engineering Research, 2012.
- [2] Ma.jabbar, Dr.prikti Chandra, B.L.Deekshatulu, cluster based association rule mining for heart attack prediction, Journal of Theoretical and Applied Information Technology, 2011.
- [3] Ms.IshtakeS.H ,Prof. Sanap S.A., "Intelligent Heart Disease Prediction System Using Data Mining Techniques", International J. of Healthcare & Biomedical Research, 2013.
- [4] Dr. K. UshaRani, analysis of heart diseases dataset using neural network approach, International Journal of Data Mining & Knowledge Management Process, 2011.
- [5] Carlos Ordonez, Edward Omiecinski, Mining Constrained Association Rules to Predict Heart Disease, IEEE. Published in International Conference on Data Mining (ICDM), p. 433-440, 2001.
- [6] NidhiBhatlaKiranJyoti, An Analysis of Heart Disease Prediction using Different Data Mining Techniques, International Journal of Engineering Research & Technology (IJERT), 2012.
- [7] ShantakumarB.Patil, Dr.Y.S. Kumaraswamy, Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction, (IJCSNS) International Journal of Computer Science and Network 228 Security , 2009.
- [8] Abhishektaneja, Heart Disease Prediction System Using Data Mining Techniques, Oriental Scientific Publishing Co., India, 2013.
- [9] M. Anbarasi, E. Anupriya, N.ch.s.n.Iyengar, Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm, International Journal of Engineering Science and Technology, 2010.
- [10] Miss. Chaitrali S. Dangare, Dr.Mrs.Sulabha S. Apte, A data mining approach for prediction of heart disease using neural networks, international journal of computer engineering and technology, 2012.
- [11] N. AdityaSundar, P. PushpaLatha, M. Rama Chandra, performance analysis of classification data mining techniques over heart diseases data base, international journal of engineering science and advanced technology, 2012.
- [12] Shadab Adam Pattekari and AsmaParveen, prediction system for heart disease using naïve bayes, International Journal of Advanced Computer and Mathematical Sciences, 2012.
- [13] LathaParthiban and R.Subramanian, Intelligent Heart Disease Prediction System using CANFIS and Genetic

- Algorithm, *International Journal of Biological and Medical Sciences*, 2008.
- [14] JesminNahar, TasadduqImama, Kevin S. Tickle, Yi-Ping Phoebe Chen, Association rule mining to detect factors which contribute to heart disease in males and females, Elsevier, 2013.
- [15] Nada Lavrac, Selected techniques for data mining in medicine, Elsevier, 1999.
- [16] TanawutTantimongcolwat, Thanakorn Naenna, Identification of ischemic heart disease via machine learning analysis on Magnetocardiograms, Elsevier, 2008.
- [17] Resul Das, Ibrahim Turkoglu, AbdulkadirSengur, Effective diagnosis of heart disease through neural networks ensembles, Elsevier, 2009.
- [18] Resul Das, Ibrahim Turkoglu, AbdulkadirSengur, Diagnosis of valvular heart disease through neural networks ensembles, Elsevier, 2009.
- [19] Oleg Yu. Atkov, Coronary heart disease diagnosis by artificial neural networks including genetic polymorphisms and clinical parameters, Elsevier, 2012.
- [20] Marcel A.J. van Gerven, Predicting carcinoid heart disease with the noisy-threshold classifier, Elsevier, 2007