Modular Approach for Heart Disease Prediction

Purushottam¹, Kanak Saxena² and Richa Sharma³

¹R. G. P. V., Bhopal - 462036, Madhya Pradesh, India; puru.mit2002@gmail.com
²S. A. T. I., Vidisha - 464001, Madhya Pradesh, India; kanak.saxena@gmail.com
³Amity University, Noida - 201313, Uttar Pradesh, India; rsharma6@amity.edu

Abstract

Objectives: In this study, a new approach based on modular approach is presented for prediction of risk level of heart disease. Methods: Prior, the utilization of computer was to fabricate a learning based clinical decision emotionally supportive network which utilizes information from therapeutic specialists and moves this information into computing device for calculations. This procedure is tedious and truly relies upon therapeutic specialist's suppositions which might be subjective. The proposed Fast Heart Disease Prediction Algorithm (FHDPA) is used to predict the severity level of heart diseases, fast and effectively. Findings: FHDPA is designed and implemented to evaluate the performance of the model. The exhibitions of the FHDPA are assessed as far as pace, arrangement correctness's and versatility and the outcomes demonstrate that the proposed FHDPA has awesome potential in foreseeing the coronary illness risk level exact and speedier. Application/Improvements: “In what capacity would we be able to transform information into valuable data that can empower social insurance specialists to settle on the successful clinical decision?” This is the principle goal of this study.

Keywords: CVD, CAD, FHDPA, Heart Disease, Modular Approach

1. Introduction

In today’s opportunity at numerous spots clinical test outcomes are regularly made in light of specialists’ instinct and experience as opposed to on the rich data accessible in numerous extensive databases. Numerous a times this procedure prompts unexpected inclinations, mistakes and a colossal therapeutic cost which influences the nature of administration gave to patients.

Today numerous clinics introduced some kind of patient’s data frameworks to deal with their social insurance or patient information.

These data frameworks normally produce a lot of information which can be in various organization such as numbers, content, graphs and pictures yet tragically, this repository that includes rich data is seldom utilized for clinical choice making. There is a considerable measure of data put away in vaults that can be utilized viably to bolster decision making in human services. This brings up a critical issue:

2. Risk Level Forecasts from Heart Disease Database

The findings of risk level from the coronary illness database are exhibited in this segment. The coronary illness repository includes the screening clinical information of heart patients. At first, the information base preprocessed to make the mining prepare more proficient.

2.1 Data Preprocessing

To avoid the non-effective and non-accurate rule set generation data preprocessing of the data is carried out.
on necessarily on the data before applying data mining algorithm. There are many steps in data preprocessing that needs to be taken for the removal of duplicate records, normalizing the values available in the database, processing for missing data points and removing unwanted data fields.

After the preprocessing step, we transform the data in the format that is suitable for mining process. The raw data is transformed into dataset that contains appropriate characteristics required for the mining process.

Furthermore, some attribute values are also transformed from continuous domain values to numeric domain value in consultation with an expert doctor.

We have designed this algorithm in two modules so that it can run faster. The algorithm for heart disease prediction is given in segment 3.

### 2.2 Dataset Description

**Source Information**

- Creators of the utilized dataset: V.A. Restorative Center, Long Beach and Cleveland Clinic Foundation: Robert Detrano, M.D., Ph.D.
- Donor: David W. Aha (aha@ics.uci.edu) (714) 856-8779.

The “num” credits tell to the vicinity of coronary illness in the patient. The scope of this characteristic is from 0 (no vicinity) to 4.

The greater part of the investigations connected with Cleveland database are centered around nonattendance (Num” esteem 0) and presence (“Num” values from 1 to 4).

Because of individual security, patient's identity data supplanted with sham values.

Number of Instances: Cleveland: 303.

The index contains a dataset related to coronary illness determination. The information was gathered from the accompanying areas:

- Cleveland Clinic Foundation (cleveland.data)

The dataset utilized as a part of this investigation contains distinctive vital qualities like ECR, cholesterol, mid-section torment and much more. The modified Attributes information and their classification based on expert’s views are given as follows:

<table>
<thead>
<tr>
<th>Total classified input Attributes: 27</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age&lt;40, Age&lt;60, Sex, CP: Typical angina, CP: Atypical angina, CP: Non-anginal pain, CP: Asymptomatic, BP&lt;80, BP&lt;90, BP&gt;90, Chol&lt;200, Chol&lt;400, Chol&gt;400, Fasting sugar&gt;120, Ecr-Normal, Ecr-Abnormal, Ecr-Probable/definite, MHR-low, MHR-Normal, MHR-Severe, Exercise induced angina, Number of major vessels:0, Number of major vessels:1, Number of major vessels:2, Number of major vessels:3, Output attribute: Class=C0, C1, C2, C3, C4</td>
</tr>
</tbody>
</table>

We have used the PART rule induction techniques without global optimization for decision making and used above modified dataset based on the expert consultation.

We have used WEKA tool to get the decision rules for classification.

### 3. Fast Heart Disease Prediction Algorithm

Algorithm: Fast Heart Disease Prediction Algorithm

**Input:** Cleveland Clinic Foundation (cleveland.data) with 10 attributes values.

**Output:** Risk level prediction of patients.

**Method:** Heart Disease Level Prediction can be done as follows.

- Scan Cleveland database (cleveland.data) for three major effective attribute values age, resting blood pressure and cholesterol against the domain range.
- If the value of these three attribute falls under normal range with respect to cardiovascular disease (CVD) parameter values, then prediction of Risk level is 0.
  - Go to exit.
- If the value of any one of these three attribute not falls under normal range with respect to cardiovascular disease (CVD) parameter values, then go to Detail Test phase.

**Detail Test phase**

- Analyze and check the report of six different attributes sex, chest pain, ecg, blockage, thalach, blood sugar against the domain range.
- Based on the domain range specified by expert doctor and Cleveland clinic foundation dataset, model predicts the risk level class from level zero to level four based on severity and then exit.

Complete domain values flow is represented in Figure 1.
3.1 Classification Decision Rules

Generated Classification Rules by the system to classify patients into heart diseases classes (C0, C1, C2, C3, C4).

**Number of Rules**: 32

- **CP**: Asymptomatic = No AND Number of major vessels: 3 = No AND Number of major vessels: 0 = Yes AND Exercise induced angina = No AND Sex = Female: C0
- **CP**: Asymptomatic = No AND Number of major vessels: 3 = No AND Age<40 = No AND Number of major vessels: 0 = Yes AND Fasting sugar>120 = Yes AND Age<60 = Yes AND CP: Typical angina = Yes AND Chol<200 = No: C0
- **Number of major vessels**: 3 = No AND Exercise induced angina = No AND Number of major vessels: 0 = Yes AND Fasting sugar>120 = Yes AND Number of major vessels: 0 = Yes AND Age<60 = Yes AND CP: Typical angina = Yes: C0
- **Number of major vessels**: 3 = No AND Exercise induced angina = No AND Number of major vessels: 2 = No AND Exercise induced angina = No AND Number of major vessels: 0 = Yes AND Age >60 = No AND CP: Typical angina = Yes AND Chol<200 = No: C1
- **Number of major vessels**: 3 = No AND Exercise induced angina = No AND Number of major vessels: 0 = Yes AND CP: Non-anginal pain = Yes AND Age<40 = No: C1
- **Number of major vessels**: 3 = No AND Exercise induced angina = No AND Number of major vessels: 0 = Yes AND CP: Typical angina = Yes AND Fasting sugar>120 = Yes AND CP: Typical angina = Yes: C1
- **Number of major vessels**: 3 = No AND Exercise induced angina = No AND Number of major vessels: 2 = No AND Exercise induced angina = No AND Number of major vessels: 0 = Yes AND Age >60 = Yes AND CP: Typical angina = Yes AND Chol<200 = No: C3
- **Number of major vessels**: 3 = No AND Exercise induced angina = No AND Number of major vessels: 1 = No AND Age<60 = Yes AND CP: Typical angina = Yes AND Chol<200 = No: C4
- **Number of major vessels**: 3 = No AND Exercise induced angina = No AND Number of major vessels: 1 = No AND Age<60 = Yes AND CP: Typical angina = Yes AND Chol<200 = No: C4
- **Number of major vessels**: 3 = No AND Exercise induced angina = No AND Number of major vessels: 1 = No AND Age<60 = Yes AND CP: Typical angina = Yes AND Chol<200 = No: C4
Modular Approach for Heart Disease Prediction

Number of major vessels: 3 = No AND Chol < 200 = No AND Sex = Female AND Age < 60 = Yes: C3

- 23 Age < 40 = No AND Fasting sugar > 120 = No AND Number of major vessels: 3 = No AND Chol < 200 = No AND Sex = Female AND Exercise induced angina = Yes: C1
- Age < 40 = No AND Fasting sugar > 120 = No AND Chol < 200 = No AND Number of major vessels: 3 = No AND Sex = Male AND Number of major vessels: 0 = Yes: C0
- Age < 40 = No AND Fasting sugar > 120 = No AND Sex = Male AND Chol < 200 = No AND Age < 60 = Yes AND CP: Atypical angina = No AND Number of major vessels: 1 = No AND Exercise induced angina = Yes: C3
- Fasting sugar > 120 = No AND Age < 40 = No AND Sex = Male AND Chol < 200 = No AND Age < 60 = Yes AND CP: Atypical angina = No AND Ecr-Normal = Yes: C1
- Fasting sugar > 120 = No AND Ecr-Normal = Yes AND Age < 40 = No AND Number of major vessels: 3 = No AND Sex = Male: C2
- Fasting sugar > 120 = No AND Ecr-Normal = Yes AND Number of major vessels: 3 = No: C0
- Fasting sugar > 120 = No AND Sex = Male AND Age < 60 = Yes AND CP: Atypical angina = No AND Chol < 200 = No AND Exercise induced angina = Yes: C1
- Fasting sugar > 120 = Yes: C3
- Chol < 200 = No AND Age < 60 = Yes AND CP: Atypical angina = No: C2
- Age < 60 = No AND Sex = Male AND Number of major vessels: 2 = No: C2

4. Performance Evaluation Summary

Correctly Classified Instances 227 75.9197 %

Incorrectly Classified Instances 72 24.0803 %

Kappa statistic 0.5985
Mean absolute error 0.1358
Root mean squared error 0.2606
Relative absolute error 52.5024 %
Root relative squared error 72.6003 %
Coverage of cases (0.95 level) 99.3311 %
Mean rel. region size (0.95 level) 47.4247 %
Total Number of Instances 299

Confusion Matrix:

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>155</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>30</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>24</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>4</td>
<td>12</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

Coverage of cases 99.3311 %

Detailed Accuracy by Class: Class wise accuracy can be found in Table 1.

5. Performance Evaluation

The trial is performed utilizing preparing information set comprises of 299 tupples with 27 unique attributes. The dataset is isolated into two sections that is 70% of the information are utilized for preparing and 30% are utilized for testing. In view of the test results appeared in Table 2, it can allude that the forecast time of FHDPA calculation is better contrasted with other calculations.

Table 2. Comparative performance evaluation

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Time Taken</th>
</tr>
</thead>
<tbody>
<tr>
<td>FHDPA</td>
<td>590 ms</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>609 ms</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>719 ms</td>
</tr>
<tr>
<td>ANN</td>
<td>1000 ms</td>
</tr>
</tbody>
</table>
6. Complexity Evaluation

As we are taking 9 attributes for heart disease class prediction. And each attribute has the different domain values as given in Table 3. So each case that has to be predicted for heart disease prediction would follow any one rule from this rule set. So it will take more time to predict the case.

Table 3. Domain range for different attributes

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Age</th>
<th>Sex</th>
<th>BP</th>
<th>Chol</th>
<th>Cp</th>
<th>Trestec</th>
<th>Ca</th>
<th>Thal</th>
<th>fbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Different</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Total condition that can be created by most of the techniques used in this domain would be

\[ 4^2 \times 2 \times 3^3 \times 4^3 \times 4^2 = 20736 \] (Twenty thousand seven hundred thirty-six).

But in modular approach in the first module, we have checked the three conditions only so the total combination would be \( 4^3 \times 3 = 36 \) only and in second module rest of the condition are checked so the total no of combination would be \( 2^4 \times 3^4 \times 4^2 = 576 \) only. So the calculated average number of conditions would be 1152 only.

7. Conclusion

The study presented a Heart disease risk level prediction using Modular Approach Algorithm for efficient decision making about the coronary illness. The programmed method can be used to produce the efficient decision-making rules. The proposed Heart disease risk level prediction utilized a modular approach on the modified dataset. At last, the experimentation was done on the UCI machine learning modified dataset. We have trained and tested the system using 10-fold method and found the accuracy up to 75.91% and 99.33 coverage of cases. The class wise detailed accuracy of the classifier is also shown in Table 1.

The data has been collected in relation to heart patients and the analysis has been done on the probability of heart attack to a person with certain medical criteria. The raw data has been processed and processed information is used in automated expert system to guide in diagnosing coronary heart disease which results in finding new and accurate rule set that relate patient data to the heart disease risk level and can provide more useful and exact information.

The purposed model proves better results as compared with some well-known classifiers given in Table 2. The proposed model can help the domain experts and even the person related with the medical field to plan for a better diagnose. Figure 1 is used to show the domain values flow diagram. The main contribution of this study is to generate accurate ruleset for heart disease risk level prediction based on the given medical parameters using modular approach.

8. References