

## Prediction Exploration for Coronary Heart Disease Aid of Machine Learning

B. Madhuravani<sup>a</sup>, Divya Priya Degala<sup>b</sup>, M. Anjaneyulu<sup>c</sup> and B.Dhanalaxmi<sup>d</sup>

<sup>a,b,c</sup>Department of Computer Science and Engineering, MLR Institute of Technology, Dundigal, Hyderabad, Telangana, India

<sup>d</sup>Department of Information Technology, Institute of Aeronautical Engineering, Dundigal, Hyderabad, India

**Article History:** Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 20 April 2021

**Abstract:** Any machine learning approach tries to map input and output while minimizing some error metrics. In addition, machine learning techniques are not classified by setting certain cut-offs or thresholds. Once again, it is not clear how to reduce this cut-off cause, but to obtain favorable metrics. The best approach is based on metrics such as precision, recall and accuracy. Proposed work on heart disease is a major concern in today's society. It is very difficult to physically control the chances of getting heart disease based on risk factors. However, monitoring machine-learning methods can be useful for estimating production from current data. This paper uses three different monitoring tools to diagnose and determine the best approach for coronary heart disease (chd) such as SVM, K-NN and ANN multilayer perceptron. In the case of chd, both controls have a temporary effect. Men who are positive for coronary heart disease tend to lower their risk factors after the onset of coronary heart disease through a blood pressure reduction trend and other programs. The experimental result is on three prediction methods like SVM, K-NN and ANN. It is to create and identify the coronary heart disease using three different supervise machine learning.

**Keywords:** Artificial Neural Networks, Coronary Heart Disease, k-Nearest Neighbour, Machine Learning, Support Vector Machine.

### 1.Introduction

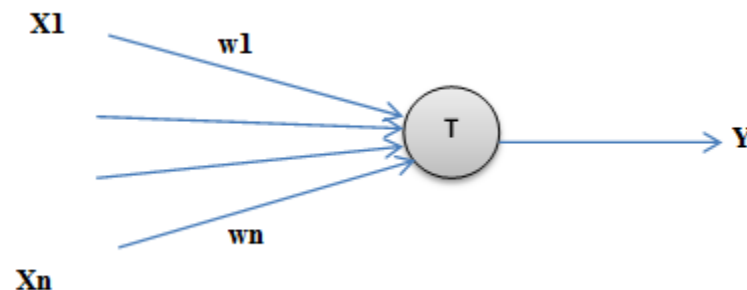
Machine learning is now a very exciting technique in data analytics. Data sources intensify and with it, the computing power increases. A simple way to access and process them is to go directly to the data and do the computation. Machine learning helps us to integrate exhaustive analytical controls by combining statistics and computer science. Data analytics, data scientists, and anyone else who wants to deal with data sources must have the expertise in them to perceive that raw data as unique trends and forecasts[8]. In this, we will learn the end-to-end process of researching data through a machine-learning lens. We can also go through machine learning algorithms; Evaluating and collecting data to evaluate the performance of machine learning. Some useful features that best represent the data are also discussed here. Machine learning is a stand in meadow of artificial intelligence that allows us to make our application productive by modifying it intelligently without the direct involvement of human programming. In other words, machine learning makes our application or system smart enough to make timely decisions using past information or from models available from data datasets stored on the system or machine [5]. Machine learning can be classified according to the knowledge signal or response of the knowledge system. Supervised machine learning is the exploration of procedures that generate general hypotheses from externally supplied contexts and then make predictions about future events. Meanwhile, it is the main objective of the supervised practice; the basic elements of the analyst's characteristics are used to form a brief replica of the circulation of the class labels. The output classification is used to allocate class labels to test cases anywhere the values of the marginal note attributes are known, but the value of the class label is not recognized.

Heart disease refers to various conditions that affect your heart. Vascular illnesses for example coronary artery disease, heart beat problems and heart defects are among the inherited heart defects in others. The term heart illness can often be replaced with the term heart illness. The illness usually refers to conditions that have narrow or blocked blood vessels leading to heart attack, chest pain or stroke. Other heart conditions that affect your heart can be well thought-out as forms of muscle, valves or beat heart illness. Heart illness is one of the leading causes of illness and is well thought-out one of the most important factors in the meadow of clinical data study in population of cardiovascular diseases. There is many data in healthcare engineering. Data mining helps in making a large collection of informational health decisions and making informed decisions [4]. This is a major concern for treating heart disease. However, it is very difficult to make out heart illness for instance diabetes, high blood pressure, high lipid, abnormal way pulse rate and many other factors. Because of such barriers, researchers have turned to modern methods such as data mining and machine learning to diagnose the disease [1].

**2. ANN Multilayer Perceptron Classifier:** Neural networks use technology that seeks to produce intelligent behavior by annoying to impersonator the structure and purpose of our nervous system [3]. This system is usually as careless as a weighted direct graph, where nodes are neurons and the edges between them are

connections between neurons. The weight at each edge indicates the type and strength of interaction between adjacent neurons.

**Perceptrons:** The pretentious kind of neural network is a perceptron, in which a single neuron has many true value or binary inputs and binary output. The inputs come through the weighted edges and are multiplied by the weights at those edges. At any given time, the net input to the neuron is the sum of all the weighted inputs. If this net limit is exceeded, then the neuron triggers and produces an output of '1', otherwise the output becomes '0'. The perceptron is exposed in figure-1.



**Figure-1: A Perceptron**

Instead, this perceptron is trained to respond to certain inputs with certain desired outputs. Initially, the weights of the perceptron are assigned randomly. Subsequently, successive inputs for perceptron can be obtained in each form  $(x_1, x_2, x_3 \dots x_n)$  during the training period. For each such input, there will be 0 or 1 desired output. The actual amount generated is strong-minded by the network input  $= w_1x_1 + w_2x_2 + \dots + w_nx_n$

If the network input is less than the threshold, the output is 0 or the output is 1. If the perceptron gives the wrong output, the following conditions may occur. The wished amount produced is 0. However, the network input is above the input. From now on, the actual output will be 1. As a result, net input decreases. In such a case, the weights should be reduced. Weight loss is determined using the perceptron learning method, for that reason, the weight loss of an upper hand should be straight relative to the input through that upper hand. Therefore,

New weight of an edge  $i = \text{old weight} - cx_i$

There are numerous approaches contingent on what C means. Currently, it is considered stable. The idea here is that if the input is too high by a certain upper hand that upper hand must have largely contributed to the error. In the future, the weight of that upper hand will decrease further. The other case where the perceptron makes a mistake when the desired output is 1, but the remaining input is less than the entry level. Thus, the net grows. In this case, the weights should be increased. The weight gain of the upper hand using the above insight should be proportional to the input through that upper hand.

From now, New weight of an edge  $i = \text{old weight} + cx_i$

where  $c$  is constant, then this process is called the fixed increment rule. The next time the same input is submitted,  $c$  can also be selected to avoid a recent error. This is called the absolute corrective rule. Behaving badly with this approach means that an input learner may not be able to remember it before learning the input.

**3. Relationships of SVM, K-NN:** Discrimination classification officially defined by an support vector machine separating hyperplane [7]. In other words, the given labeled training data provide an appropriate hyperplane that categorizes the practice and approach being monitored. In sample credit, k-neighboring approach is a non-parametric technique used for classification and regression [9][12]. In both cases, the input feature contains  $k$  closet training instances in place. Output in the K-NN classification is a class association [11]. Artificial neural networks are vaguely induced computing systems by two-word neural networks that establish physical brains.

**4. Experimental Result:** Experimental Outcome is the best approach is to make out and identify coronary heart disease using three different supervisor machine-learning approaches: support vector machine, K nearest neighbor and ANN multilayer perceptron [6][13]. In this regard, Bhimavaram Hospital in Bhimavaram is a reflective model of men at high risk of heart disease. The CHD case has two controls unequally. Many women who are CHD positive have undergone blood pressure reduction therapy and other programs to reduce their risk factors after the onset of CHD. In some cases, measurements were made after these treatments. This data is taken from a large dataset [2].

| S.No | Age | Systolic Blood | Tobacco | LDL | adiposity | famhist | type | obesity | Alcohol | CHD |
|------|-----|----------------|---------|-----|-----------|---------|------|---------|---------|-----|
| .    | e   |                | o       | L   | y         | t       | e    | y       | l       | D   |

|   |    | Pressur<br>e |       |      |       |   |    |       |       |   |
|---|----|--------------|-------|------|-------|---|----|-------|-------|---|
| 1 | 38 | 160          | 12.00 | 5.73 | 23.11 | 1 | 49 | 25.30 | 97.20 | 2 |
| 2 | 35 | 144          | 0.01  | 4.41 | 28.61 | 2 | 55 | 28.87 | 2.06  | 2 |
| 3 | 37 | 118          | 0.08  | 3.48 | 32.28 | 1 | 52 | 29.14 | 3.81  | 1 |
| 4 | 39 | 170          | 7.50  | 6.41 | 38.03 | 1 | 51 | 31.99 | 24.26 | 2 |
| 5 | 36 | 134          | 13.60 | 3.50 | 27.78 | 1 | 60 | 25.99 | 57.34 | 2 |

This Table-1 determines the raw data, there are no missing values in our data and to equalize the values of different attributes and apply feature scaling and min-max scaling. Additionally, the data now looks like the following

| S.No<br>. | Ag<br>e | Systolic<br>Blood<br>Pressur<br>e | Tobacc<br>o | LD<br>L | adiposit<br>y | famhis<br>t | typ<br>e | obesit<br>y | Alcohol | CH<br>D |
|-----------|---------|-----------------------------------|-------------|---------|---------------|-------------|----------|-------------|---------|---------|
| 1         | 38      | 160                               | 12.0        | 5.73    | 23.11         | 0           | 49       | 25.30       | 97.20   | 1       |
| 2         | 35      | 144                               | 0.01        | 4.41    | 28.61         | 1           | 55       | 28.87       | 2.06    | 1       |
| 3         | 37      | 118                               | 0.08        | 3.48    | 32.28         | 0           | 52       | 29.14       | 3.81    | 0       |
| 4         | 39      | 170                               | 7.50        | 6.41    | 38.03         | 0           | 51       | 31.99       | 24.26   | 1       |
| 5         | 36      | 134                               | 13.60       | 3.50    | 27.78         | 0           | 60       | 25.99       | 57.34   | 1       |

As well as, the Table-2 data description after pre-processing is as follows:

| S.N<br>o. | Ag<br>e | Systolic<br>Blood<br>Pressur<br>e | Tobacc<br>o | LDL   | Adiposit<br>y | Famhis<br>t | Type  | Obesit<br>y | Alcohol | CHD    |
|-----------|---------|-----------------------------------|-------------|-------|---------------|-------------|-------|-------------|---------|--------|
| Count     | 38      | 462.0                             | 462.0       | 462.0 | 462.0         | 462.0       | 462.0 | 462.0       | 462.0   | 462.00 |
| Mean      | 35      | 31.9                              | 3.6         | 4.7   | 25.4          | 0.58        | 53.10 | 26.04       | 17.05   | 0.346  |
| STD       | 37      | 17.5                              | 4.6         | 2.07  | 7.8           | 0.49        | 9.82  | 4.21        | 24.48   | 0.476  |
| Min       | 39      | 0.0                               | 0.0         | 0.98  | 6.74          | 0.00        | 13.00 | 14.70       | 0.00    | 0.00   |
| 25%       | 36      | 19.6                              | 0.05        | 3.28  | 19.77         | 0.00        | 47.00 | 22.98       | 0.51    | 0.00   |
| 50%       | 38      | 28.2                              | 2.0         | 4.34  | 26.11         | 1.00        | 53.00 | 25.80       | 7.51    | 0.00   |
| 75%       | 35      | 40.17                             | 5.50        | 5.79  | 31.23         | 1.00        | 60.0  | 28.50       | 23.89   | 1.00   |
| Max       | 37      | 100.00                            | 31.20       | 15.33 | 42.49         | 1.00        | 78.0  | 46.58       | 147.19  | 1.00   |

Table-3: Table shown Different Topographies

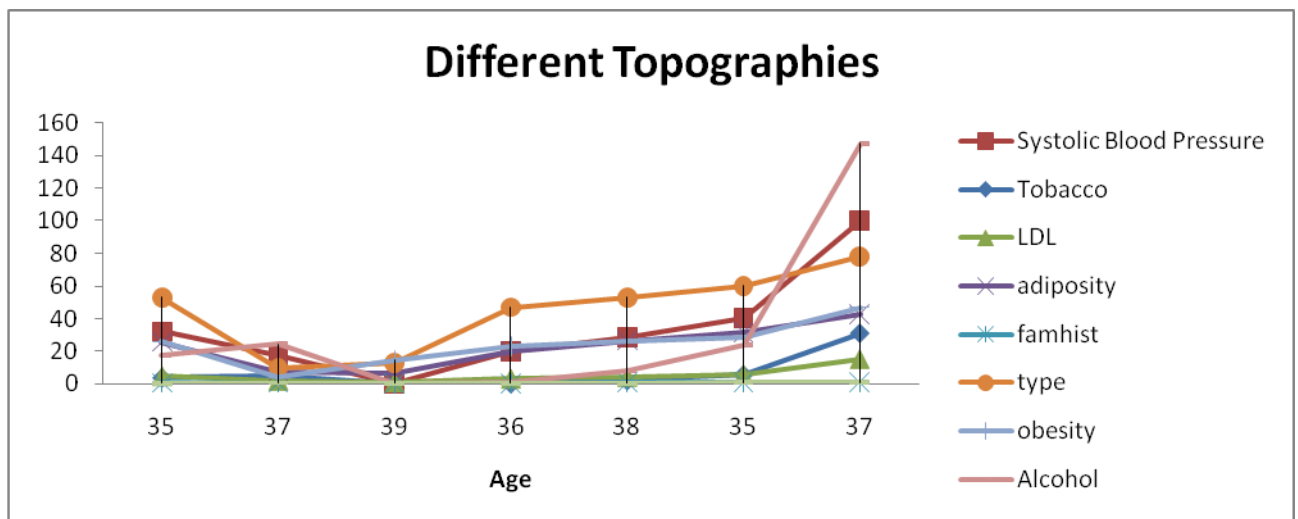


Figure-2: Different Topographies

Afterward intrigue different features of each other; we get some useful results as follows. The above scatter graph gives us information about the balance between age and age; we can observe groups before 20 years, after 40 years and mostly in 50 years.

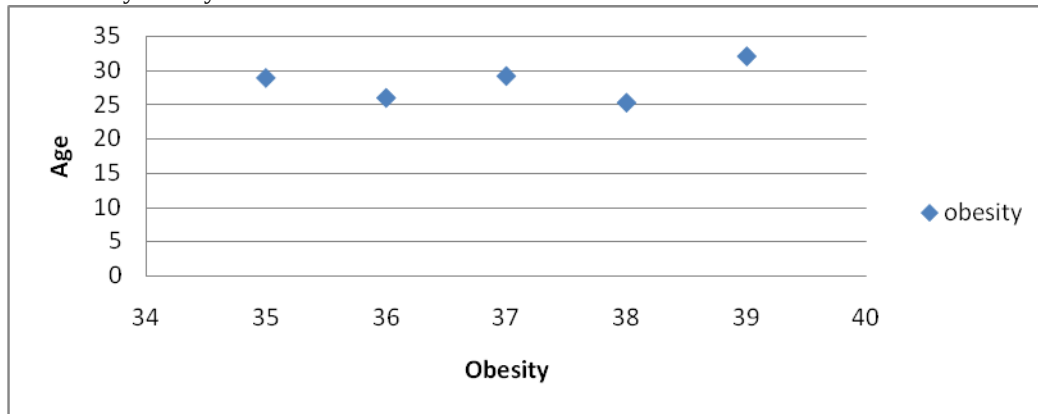


Figure-2:scatter graph gives obesity vs age

In tobacco versus age graph intake, we can see that intake increases after 30 years. It was observed that the maximum alcohol intake begins after the age of 25 according to the alcohol and age graph intake.

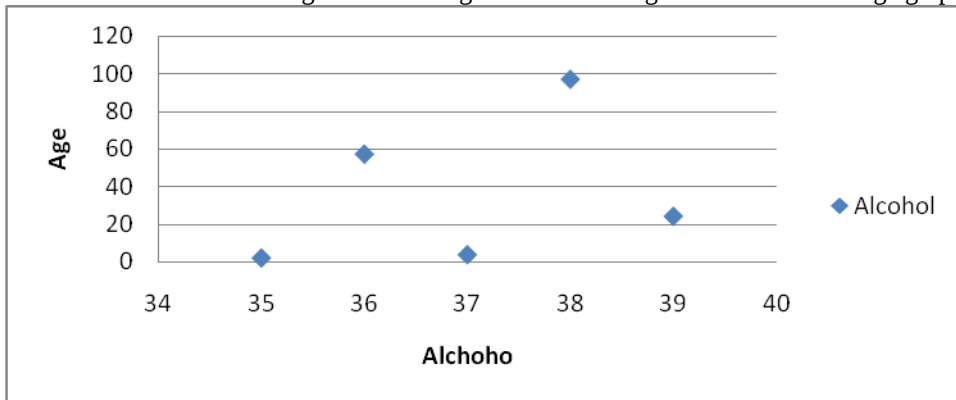


Figure-3: Alcohol vs. Age Graph it was observed that maximum alcohol intake begins after the age of 25 years.

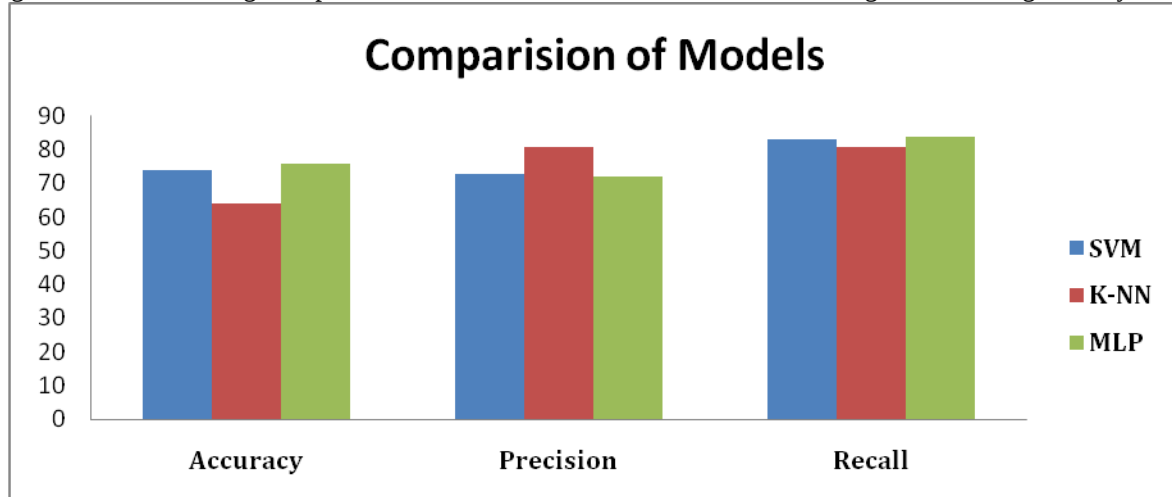


Figure-4: Comparision of Three not the same Models

After applying these three prediction methods SVM, K-NN and ANN, the following results obtained with the best accuracy in ANN, the best accuracy in K-NN, and the best recall in ANN. Therefore, the K-near neighbors and support vector machine is the best approach using an artificial neural network of these three.

**6. Conclusion:**Heart disease refers to many conditions that affect your heart. Disease under the umbrella of heart disease includes vascular disease such as coronary heart disease. Diagnosis of the disease is considered one of the most important aspects in the field of clinical data analysis. Several methods have been applied to assessment methods, with the best accuracy obtained in ANN, the best precision in K-NN and the best recall in ANN.

Therefore, the K-nearest neighbors and support vector machine is the best approach out of these three using an artificial neural network.

## REFERENCES

1. **Enke, David, and Suraphan Thawornwong:** The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with Applications* 29: 927–40, 2005. [CrossRef]
2. **Hamid Saadatfar, Samiyeh Khosravi, Javad Hassannataj Joloudari, Amir Mosavi and Shahaboddin Shamshirband:** A New K-Nearest Neighbors Classifier for Big Data Based on Efficient Data Pruning, *Mathematics* 2020, 8, 286; doi:10.3390/math8020286, Received: 26 November 2019; Accepted: 8 January 2020; Published: 20 February 2020.
3. **Jakub Horak, Jaromir Vrbka and Petr Suler:** Support Vector Machine Methods and Artificial Neural Networks Used for the Development of Bankruptcy Prediction Models and their Comparison, *Journal of Risk and Financial Management*, Received: 29 January 2020, accepted: 23 March 2020, published: 24 March 2020, 13, 60, DOI: 10.3390/jrfm13030060.
4. **Jones LD, Golan D, Hanna SA, Ramachandran M:** Artificial intelligence, machine learning and the evolution of healthcare: a bright future or cause for concern? *Bone Jt Res.* 2018, 7: 223–225, doi:10.1302/2046-3758.73.BJR-2017-0147.R1. [CrossRef]
5. **Meenakshi:** Machine Learning Algorithms and Their Real-Life Applications: A Survey Meenakshi Research Scholar, International Conference on Innovative Computing and Communication (ICICC 2020) U.I.E.T (M.D.University), Rohtak-124001, India.
6. **Najat Ali, Daniel Neagu, and Paul Trundle:** Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets, Received: 13 June 2019, Accepted: 24 September 2019, Published online: 6 November 2019 © The Author(s) 2019, A Springer Nature Journal, *SN Applied Sciences* (2019) 1:1559, <https://doi.org/10.1007/s42452-019-1356-9>.
7. **P. T. Noi and M. Kappas:** Comparison of random forest, K-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery, *Sensors*, vol. 18, no. 1, p. 18, 2018.
8. **Rene Y. Choi; Aaron S. Coyner; Jayashree Kalpathy-Cramer; Michael F. Chiang; J. Peter Campbell:** Introduction to Machine Learning, Neural Networks, and Deep Learning, *Translational Vision Science & Technology* February 2020, Vol.9, 14. DOI:<https://doi.org/10.1167/tvst.9.2.14>.
9. **S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang:** Efficient kNN classification with different numbers of nearest neighbors, *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1774–1785, May 2018.
10. **Wenchao Xing and Yilin Bei:** Medical Health Big Data Classification Based on KNN Classification Algorithm, *IEEE ACCESS*, Digital Object Identifier 10.1109/ACCESS.2019.2955754, Received November 4, 2019, accepted November 14, 2019, date of publication November 25, 2019, date of current version February 17, 2020.
11. **Z. Chen:** Identification of Android malicious behaviors based on k nearest neighbor algorithm and least squares support vector machine, *J. Jilin Univ.*, vol. 53, no. 4, pp. 720–724, 2015.
12. K. Sai Prasad, Dr. S Pasupathy, "Deep Learning Concepts and Libraries Used in Image Analysis and Classification", *TEST Engineering & Management*, Volume 82, ISSN: 0193 - 4120 Page No. 7907 – 7913.
13. N.Thulasi Chitra, K.Pushpa Rani, Roja, "A Credit Card Fake Detection System using Image Cryptography" *International Journal of Recent Technology and Engineering (IJRTE)* with ISSN: 2277-3878 (Online), Volume-7 Issue-6, March 2019