*International Journal of Engineering & Technology*, 7 (4.6) (2018) 138-140

International Journal of Engineering & Technology

*Website: www.sciencepubco.com/index.php/IJET*

***Research paper***

# Co- Disease Prediction using Multileyer Perceptron and Classification from Diabetic Medical Data Sets

**Shahebaz Ahmed Khan[1],Dr.Ramakrishna Seemakurthi[2],Dr Akhil Jabbar[3]**

*[1]PhD Research Scholar, Dept of CSE, JJT University*
*[2]PhD Supervisor, JJTU and Principal, Swami Vivekananda Institute of Technology, Hyd*
*[3]Professor, Vardhaman College og Engineering*
**Corresponding author E-mail: sakklkyr@gmail.com*

**Abstract**

Artificial Neural Networks (ANN) techniques have the important concepts those can be used in the present scenario of the medical world. It has made the medical field to formulate easy steps to detect and predict the diseases like diabetes, thrombocytopenia, heart diseases, brain tumor, cancer etc. The classification methods available in the data mining theories and ANN gradually help to predict the data for the future analysis by building the classification models. In this paper, the results and the research work carried out on diabetic medical data using the artificial neural network algorithms like multilevel perception and its application over such data so as to predict the diseases are discussed. The rules developed will be helpful to detect the co-disease in the diabetic patients and we have ranked them as per the final classifier for prediction. The proposed classification algorithm has accurately predicted the data with and without feature subset selection.

***Keywords***: *ANN, classification methods, final classifier ,co-disease and multilevel perceptron*

## 1. Introduction

Data mining is the process of discovering the patterns hidden and the extraction of these patterns from the various sources so as to analyze the data and then opting to decision making. Today the roots and branches of data mining have evaluated many theories with its application on various data sets. Based on the two learning methods[5] which are available in mining concepts we perform the different data mining tasks like classification, association rule and pattern recognition are being performed etc. In supervised learning, the class is predetermined and it is known in the data set or training data set. By this, we will search for the output pattern. Artificial Neural Network is one of the most effective and robust concepts[12] where the supervised learning can be applicable for accurate results. Many prediction based works can be easily analyzed where the ANN is used. The classification model used in our work easily explains the nature of the data for the feature predication. The multilevel perception algorithm and the naive based algorithm both have been used in the current work so as to predict the co-diseases in the diabetic patients. The results obtained can be deeply analyzed in the medical area which contains vast information that relates to different diseases[8]. Here, the data that may can be used to detect the other diseases from the existing problems of the patients.

Though there are different data mining techniques and ANN algorithms that can be used to find out the future diseases that shall arise. In this work to predict the future diseases and the symptoms, the classification model like MLP[1] has been used for the prediction of the co-disease in the diabetic patients. Co-disease is the disease which shall occur in a person along with the already present disease. Here, the co-disease prediction is formulated both from the syndromes and other diseases. Classification plays a major role in the diagnosis of diseases and data classification is the categorization of data so that it can be used efficiently and effectively Based on training set, we can predict the categorical class labels and classifies data.

## 2. Artificial Neural Networks And Multilayer Perceptron

With the idea of biological network of human brain functioning artificial neural network has enveloped. These are the computing systems that constitute the programming to solve the complex tasks with its inter connected networks based on the deep learning programs. It contains the connected units and these units are called artificial neurons. Highly interconnected processing elements are present which will process the information to the external states with its computing capability.

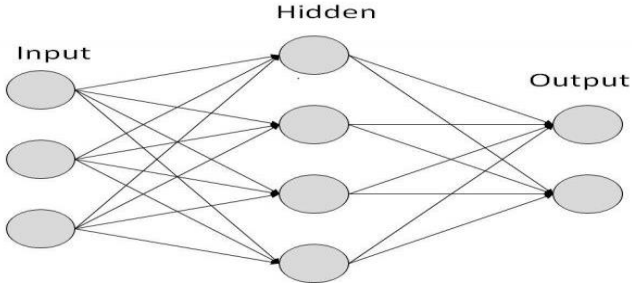



**Figure 2.1**-ANN Model

Based on the types the artificial neural networks can be unidirectional and multidirectional which contains loops

**Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.**

possibility. The former one is called feed forward ANN and later one is called feedback ANN.
ANN are capable of learning and they need training for these learning strategies. It follows supervised, unsupervised and reinforced learning strategies. Bayesian networks are the examples where the ANN is implemented with efficacy [6].

### 2.1 Feed Forward ANN

Multilayer perceptron is the best example of the Feed forward [7] ANN. MLP uses the supervised learning and it consists of three layers of nodes and these layers are input, output and hidden layer. The supervised learning method used here in training the data technique is called back propagation. Even the algorithm of MLP has the capability to distinguish the non linearly separable data. Multilayer perceptron does not refer to a single perceptron that has multiple layers. But, it contains many perceptrons that are organized well into layers. In this we can use mathematical models with classification.

## 3. Co-Diseases And Diabetes In Recent Scenario

Diabetes is a syndrome and not a disease in which the blood glucose or sugar levels are too high. A human body can have type 1 or type 2 diabetes. Having too much glucose in our body can cause serious problems. It can damage the eyes, kidneys, nerves heart attack can occur by causing the severe heart pain. More over it is a metabolic disorder where polyurea, thirsty and hunger are common symptoms[4]. Approximately, 90 percent of the people suffer with type 2 diabetes and 10 % with type 1 diabetes.

Type 1 usually occurs in the persons below 40 years and type 2 at any age[10]. It causes the path to follow our body with other diseases. The complications with diabetes are very severe and can cause mental health issues, eye complications, gum diseases, neuropathy, thrombocytopenia, heart pain and brain stroke etc. So, it becomes very important to predict the diseases that occur with diabetes at early stage itself so as to avoid the future health problems. In 2013,it was estimated that there are 382 million people with diabetes in the world.

## 4. Related Work

Large and various research works are carried out for medical diagnosis of various diseases. In our proposed approach we have attempted to predict possibility of heart disease in diabetic patients using feature subset selection and classification. It was closely diagnosed with future subset selection and without future subset selection. A feature or attribute or a variable refers to an aspects of the data. Features can be discrete, continuous, or nominal. Generally features are of three types 1) Relevant 2) Irrelevant 3) Redundant. Feature subset selection has been an active and fruitful field of research and development for decades in pattern recognition, machine learning and data mining. It has proven in both theory and practice effective in enhancing learning efficiency, increasing predictive accuracy and reducing complexity of learned results.

M.A.Jabbar et al[3] has proposed heart disease prediction by Knowledge discovery using associative classification. Cluster based association rule mining for heart disease prediction was proposed in [4]. The prediction of Heart disease, Blood Pressure and Sugar with the aid of neural networks was proposed by Niti Guru et al. [9]. Sellappan Palanniappan has proposed Intelligent Heart Disease Prediction System by using Naïve bayes and neural networks[11].

## 5. Predication Of Disease And Proposed Work

In the proposed method the data has been classified for prediction of disease by applying feature subset selection. The rules generated are considered as the classifiers. The proposed approach prunes irrelevant, redundant attributes and generates compact rule set. Accuracy of our method improves over other classification algorithms. In order to predict the co-diseases that come with the diabetes, our work has diagnosed the case sheets of diabetic patients. The attributes selected were age, gender, various cholesterol levels in our body, triglycerides, ldl and hdl cholesterols etc. There was a change in the levels of the confidence and accuracy. The instances selected were of good sample and almost for every 100 instances it was predicted that 19-20 are falling in the category of the heart diseases. It was pre conditional that

if
diabetes = ‘yes’ and age>=45,
gender=male||female,
total cholesterol >230,
HDL<35,
LD L >130,
VLDL >30,
TTL >5.0,
Triglycerides >190
then
disease=’yes’
else
disease=’no’

## 6. Results

The results obtained are as follows in table no.6.1 and table no.6.2.The results clearly depict the changes in the accuracy with and without the feature subset selection. The sets of input data have been mapped to the appropriate output data. The results show how close the predictions are to the outcomes. Based on the correctly and incorrectly classified instances the results are framed in the tables.

**Table 6.1** Accuracy Measure

| Without Future Subset Selection | With Future Subset Selection |
|---|---|
| **91.26** | **93.20** |

The changes in the levels of accuracy after the attribute removal are as follows in table no 5.2.The results are added after removing the attributes like age and gender .the change in the accuracy was approximately 0.93 to 1.94.

**Table 6.2**

| | |
|---|---|
| Before F.S | **91.26** |
| After removing age | **92.23** |
| After removing gender | **93.20** |

With future subset Chi-Square algorithm the accuracy was 93.2 and the confusion matrix built was as follows

| **a** | **b** |
|---|---|
| 73 | 2 |
| 5 | 23 |

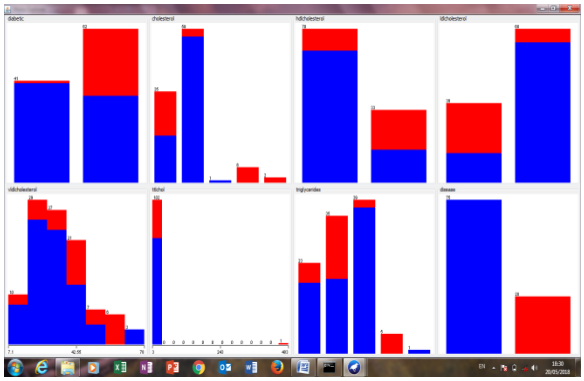


**Figure 6.1**-Visualization of the attributes (excluding age and gender)

Data visualization helps to use and understand the more complex data easily and the decision makers can effectively predict the data outcomes by data visualization.

## 7. Conclusion

From our work, it can be concluded that we have predicted the most occurring problem like heart disease in diabetic patients .In this aspect, the classification techniques applied with and without future subset selection shows the accuracy of the algorithm used. This can be a future application for the medical field for the early prediction of the diseases in the diabetic patients. The chance of occurrence of co disease can be estimated and treated with at most care with the implementation of the proposed work. The possibility of the heart diseases in diabetic patients can be avoided by the pre care and good health maintenance.

## References


[1] Hastie, Trevor. Tibshirani, Robert. Friedman, Jerome. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York, NY, 2009.

[2] M.A.Jabbar, B.L.Deekshatulu, Priti Chandra," anevolutionary algorithm for heart disease prediction,CCIS pp 378-389 Springer Verlag 2012.

[3] M.Ambarasi et al,"Enhanced prediction of heart diseases with feature subset selection using genetic algorithm"IJEST Vol 2(10) pp 5370-5376(2010)

[4] Symptoms of low bold sugar webMD from original on 18 june 2016,Retrieved 29 June 2016.

[5] Rochester, N.; J.H. Holland; L.H. Habit; W.L. Duda (1956). "Tests on a cell assembly theory of the action of the brain, using a large digital computer". IRE Transactions on Information Theory. 2 (3): 80–93. doi:10.1109/TIT.1956.1056810

[6] Ciresan, Dan; Giusti, Alessandro; Gambardella, Luca M.; Schmidhuber, Juergen (2012). Pereira, F.; Burges, C. J. C.; Bottou, L.; Weinberger, K. Q., eds. Advances in Neural Information Processing Systems 25 (PDF). Curran Associates, Inc. pp. 2843–2851.

[7] Chen, Bo; Polatkan, Gungor (2011). "The Hierarchical Beta Process for Convolutional Factor Analysis and Deep Learning" (PDF). Machine Learning .

[8] M.A.Jabbar, B.L.Deekshatulu, Priti Chandra,
"Cluster based association rule mining for heart diseaseprediction, JATIT, Vol 32 no 2 October (2011).

[9] Niti Guru, Anil Dahiya, Navin Rajpal, "Decision Support System for Heart Disease Diagnosis Using Neural Network", Delhi Business Review, Vol. 8, No. 1 (January - June 2007).

[10] Definition and diagnosis of Diabetes Millitus and Intermediate Hyperglycemia, report of WHO/IDE,2006 P:21 ISBN 978-92-4-159493-6

[11] Sellappan Palaniappan, Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", IJCSNS International Journal of Computer Science and Network Security, Vol.8 No.8, August 2008

[12] Da, Y.; Xiurun, G. (July 2005). T. Villmann, ed. An improved PSO-based ANN with simulated annealing technique. New Aspects in Neurocomputing: 11th European Symposium on Artificial Neural Networks. Elsevier. doi:10.1016/j.neucom.2004.07.002Education and Counseling 53, 309–313.