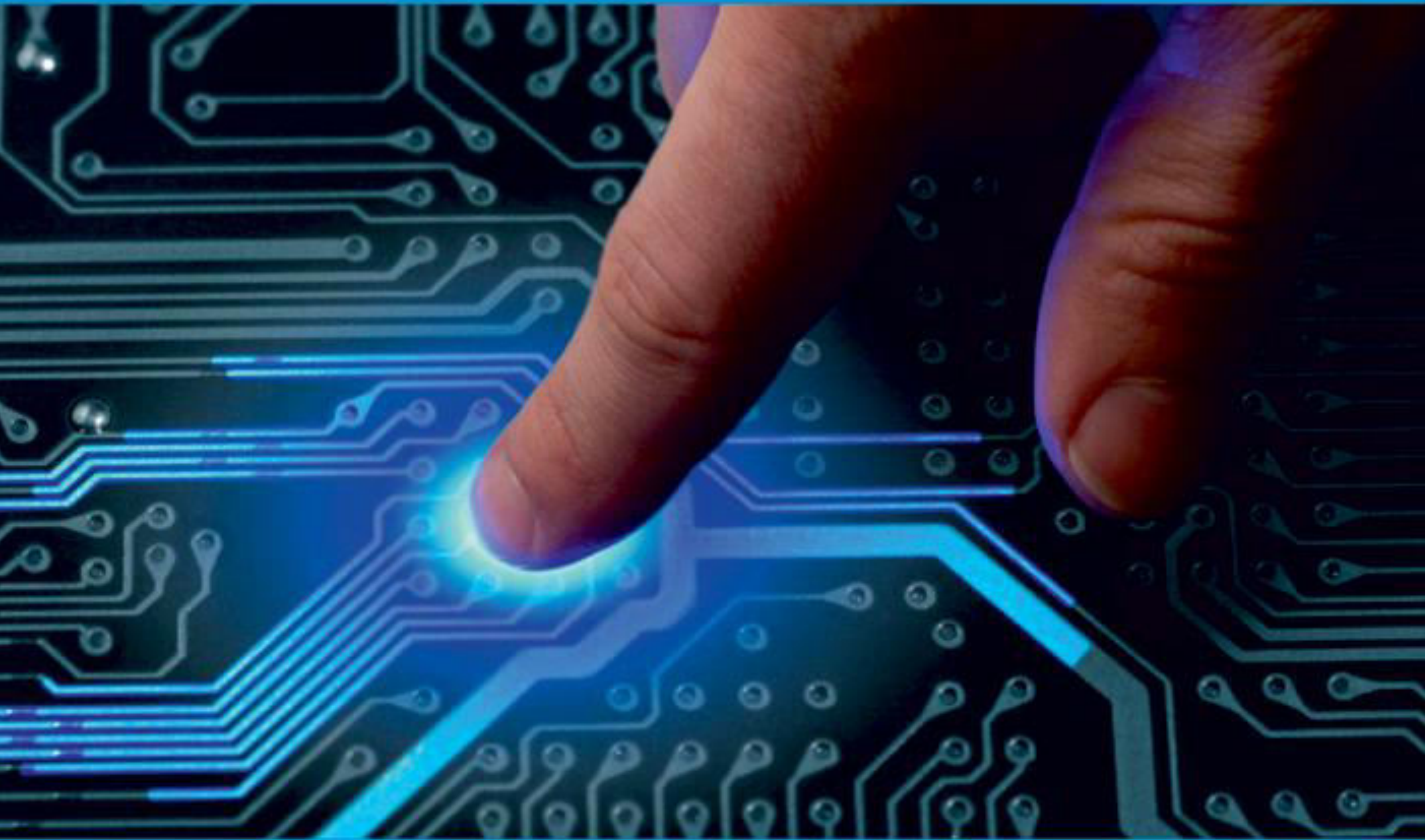
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 12, December 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 12, December 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1112022 |

Predicting Heart Wellness: A Survey on Machine Learning Approach

Ashiq Ahammed [1], Safa Akbar [2] , Sahla Abdul Latheef [3], Sidhartha VS [4], Santhi P [5]

B. Tech Final Year Student, Department of Computer Science and Engineering,IES College of Engineering, Chittilappilly, Thrissur, Kerala, India[1]

B. Tech Final Year Student, Department of Computer Science and Engineering,IES College of Engineering, Chittilappilly, Thrissur, Kerala, India[2]

B. Tech Final Year Student, Department of Computer Science and Engineering,IES College of Engineering, Chittilappilly, Thrissur, Kerala, India[3]

B. Tech Final Year Student, Department of Computer Science and Engineering,IES College of Engineering, Chittilappilly, Thrissur, Kerala, India[4]

Assistant Professor, Department of Computer Science and Engineering,IES College of Engineering, Chittilappilly, Thrissur, Kerala, India[5]

ABSTRACT:The surge in heart diseases among individuals at a young age is becoming increasingly apparent, marking a concerning trend. This escalation can be primarily attributed to the profound impact of lifestyle choices on cardiovascular health. Sedentary habits, poor dietary practices, and heightened stress levels prevalent in contemporary living contribute significantly to this rise. Recognizing the relevance of lifestyle in this health issue is crucial as it highlights the need for preventive measures. Early prediction of the possibility of heart disease emerges as a pivotal strategy in addressing this challenge. Detecting potential risks at an early stage allows for timely interventions, potentially decreasing mortality rates and enhancing the life span of patients. In this context, the focus of the study is directed towards exploring and developing effective methods for early prediction, emphasizing the importance of proactive healthcare strategies in mitigating the growing burden of heart diseases in younger populations.

I. INTRODUCTION

The global increase in heart diseases over the years has emerged as a significant public health challenge. According to the latest World Health Organization (WHO) survey, there has been a notable uptick in the prevalence of cardiovascular issues across diverse regions and age groups. Lifestyle factors, including sedentary behavior, unhealthy diets, and rising stress levels, have been identified as key contributors to this concerning trend. The WHO survey underscores the urgent need for early detection of heart diseases to address the growing burden on healthcare systems worldwide.

The survey findings reveal a compelling link between lifestyle choices and the rise in heart diseases, emphasizing the importance of proactive healthcare measures. Early detection becomes paramount in mitigating the impact of cardiovascular issues, enabling timely interventions and personalized health management. The WHO survey serves as a crucial tool in understanding the global landscape of heart diseases, emphasizing the need for comprehensive strategies that focus on early detection to curb the escalating rates and enhance the overall cardiovascular health of populations worldwide.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 12, December 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1112022 |

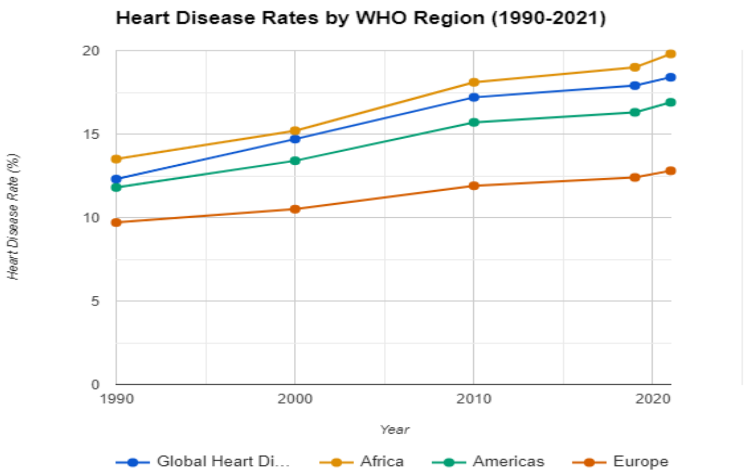


Figure.1 Heart Disease Rate by WHO

The graph depicts the increase in heart disease rate over time.There is serious that can predict accurately,if someone might get heart disease. With more people facing heart issues, especially because of how we live. Early detection means we can take steps to prevent or manage heart diseases sooner, making a big difference in people's health. As heart diseases become more common, having tools for prediction is super important to improve overall health outcomes.The latest AI and machine learning technologies offer new approaches for reliable and accurate heart disease prediction. This study explores various methodologies in addressing this advancement.

II. LITERATURE REVIEW

The literature review presented here encapsulates a comprehensive overview of research endeavors focused on the utilization of data mining techniques for the detection and prediction of heart disease. The exploration is marked by diverse methodologies and approaches adopted by various researchers in their quest to enhance diagnostic accuracy and prognostic capabilities in the realm of cardiovascularhealth.

Jyoti Soni et al. [1] initiated their investigation by employing data mining techniques, including Naive Bayes, Decision List, and KNN, to discern patterns associated with heart disease. Notably, their findings underscored a limitation in the performance of Classification based on clustering.

P.K Anooj et al. [2] proposed a distinctive weighted fuzzy rule-based system, introducing an automated approach for generating weighted fuzzy rules and subsequently developing a fuzzy rule-based decision support system. The integration of Mamdani fuzzy inference system in the prediction of heart disease marked a significant contribution.

Nidhi Bhatla et al. [3] delved into an analysis of diverse data mining techniques for heart disease prediction. Their observations highlighted the superiority of neural networks with 15 attributes and the efficacy of decision trees when coupled with genetic algorithms and feature subset selection.

Aditya Methaila et al. [4] expanded the repertoire of classification modeling techniques by incorporating Decision Trees, Naïve Bayes, Neural Networks, andweighted association algorithms (Apriori and MAFIA) in their approach to heart disease prediction.Shimpy Goyal et al. [5] engaged in a discussion on data mining techniques, specifically K-means and Apriori algorithms, to predict heart disease. Their work not only presented predictive methodologies but also addressed challenges inherent in disease detection and diagnosis.

M Akhil Jabbar et al. [7] introduced an innovative approach, utilizing anassociative classification-based genetic algorithm for heart disease prediction. Their rationale emphasized the efficiency gained by employing genetic algorithms to optimize attribute sets, thereby mitigating complexities associated with large datasets.

R Sethukkarasi et al. [8] contributed to the literature with novel neuro-fuzzy techniques pre-processed by genetic algorithms. Their approach incorporated a four- layered fuzzy neural network and a radial basic function neural network, offering a unique perspective on predictive modeling for heart disease.

Mohammed Abdul Khaleel et al. [9] introduced a method employing the Apriori data mining technique for disease prediction. The incorporation of graphical representations and the development of a prototype showcased the practical applicability of their proposed method.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 12, December 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1112022 |

Boshra Bahrami et al. [10] systematically evaluated different classification techniques, including Decision Trees, KNN, and Naive Bayes, for heart disease diagnosis. Their performance evaluation positioned Decision Trees as the optimal classifier for heart disease diagnosis within their dataset.

Deepika N et al. [11] introduced the concept of Pruning Classification Association Rule (PCAR), derived from the Apriori algorithm, for heart disease prediction. Their method, involving the deletion of minimum frequency items from item sets, aimed at refining the discovery of frequent item sets.

Collectively, these studies underscore the breadth and depth of research efforts directed towards leveraging data mining techniques in the domain of heart disease prediction. The amalgamation of traditional machine learning algorithms, fuzzy logic, genetic algorithms, and association rule mining reflects the interdisciplinary nature of this field, with each study contributing nuanced insights and methodologies to advance the understanding and application of data mining in cardiovascular health.

III. BACKGROUND KNOWLEDGE

A. MACHINE LEARNING

Machine learning is a branch of artificial intelligence that empowers computers to learn and improve performance without explicit programming. It revolves around creating algorithms that enable systems to recognize patterns and make predictions based on data. This transformative technology has wide-ranging applications, from healthcare to finance. In healthcare, for example, machine learning aids in predicting and diagnosing conditions like heart disease by analyzing large datasets. Its capacity to automate tasks and enhance performance over time makes it a valuable tool, driving innovation and revolutionizing problem-solving approaches in various fields. As technology advances, the influence of machine learning is set to expand further, shaping our data-driven future.

B. SUPERVISED LEARNING

Supervised learning is a fundamental concept in machine learning where computers are trained using labeled data to make predictions or classifications. In this process, the algorithm is provided with input-output pairs, allowing it to learn patterns and relationships. The "supervisor" guides the model by providing correct answers during training, enabling the algorithm to make accurate predictions when faced with new, unseen data. This approach is widely used in tasks such as image recognition, speech processing, and recommendation systems, making supervised learning a crucial and accessible tool for solving a variety of real-world problems.

C. UNSUPERVISED LEARNING

Unsupervised learning is a type of machine learning where computers learn without being explicitly told what to look for. In this approach, the system explores data on its own, finding patterns and relationships without predefined labels or guidance. It's like letting the computer discover insights independently. Unsupervised learning is often used for tasks like clustering similar data points or reducing the dimensionality of complex datasets. This flexible and exploratory nature makes unsupervised learning a powerful tool for uncovering hidden structures and trends within information, contributing to a deeper understanding of data in various applications.

Machine learning offers various algorithms for both classification and prediction tasks, catering to different data characteristics and problem contexts. For early geart disease detection ,algorithm with respect to prediction has to be implemented.

IV. CLASSIFICATION TECHNIQUES USED IN RELATED WORKS

NAIVE BAYES

Naive Bayes is a probabilistic machine learning algorithm used for prediction tasks, particularly in text classification and spam filtering. It relies on Bayes' theorem to calculate the probability of a certain event based on prior knowledge of related events[4]. Despite its simplicity and assumption of feature independence (naivety), Naive Bayes often performs well and is computationally efficient, making it a popular choice for predictive modeling in various applications.

DECISION TREE

Decision Tree is a popular machine learning algorithm used for prediction tasks. It creates a tree-like model that recursively splits the data based on the most significant features, leading to a series of decision nodes[10]. By traversing the tree, predictions are made for new instances based on their feature values. Decision Trees are versatile,

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 12, December 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1112022 |

interpretable, and effective for both classification and regression tasks, making them valuable tools in predictive modeling.

K-NEAREST NEIGHBORS

L-Nearest Neighbors (KNN) is a simple yet effective machine learning algorithm for prediction tasks. It classifies or predicts an outcome based on the majority class or average of its nearest neighbors in the feature space. With its intuitive approach, KNN is particularly suitable for scenarios where the underlying data exhibits local patterns or clusters. Its versatility and ease of implementation make it a popular choice in various fields, such as image recognition and recommendation systems[1].

APRIORI

Apriori is an algorithm commonly used in association rule mining, particularly for predicting patterns in transactional datasets. It identifies frequent itemsets, discovering relationships and dependencies among items. By establishing associations between items that often co-occur[5], Apriori enables predictions about future item occurrences in similar contexts. Widely applied in market basket analysis, it helps businesses anticipate customer preferences and optimize product recommendations based on historical transaction data.

MAFIA

In machine learning, the term "Mafia Technique" refers to the practice of intentionally introducing noise or misleading information into a dataset during the training phase. This technique aims to assess the robustness of a predictive model by challenging it with deceptive elements[4]. By incorporating elements of uncertainty and unpredictability, the model is forced to become more resilient, enhancing its ability to make accurate predictions in real-world scenarios where noise and misinformation are prevalent. The Mafia Technique serves as a strategic tool for fortifying machine learning models against potential pitfalls and improving their overall performance.

FUZZY RULE-BASED DECISION SUPPORT SYSTEM

Fuzzy Rule-Based Decision Support Systems play a pivotal role in prediction tasks by incorporating fuzzy logic to handle uncertainty and imprecision in data. These systems use a set of rules defined by human experts, allowing for the expression of vague and subjective knowledge[2]. By considering degrees of membership rather than binary conditions, Fuzzy rule provide more nuanced predictions, making them particularly valuable in complex, real-world scenarios where traditional approaches may fall short. With their ability to model uncertainties, these systems contribute to enhanced decision-making and prediction accuracy in diverse fields.

PRUNING CLASSIFICATION ASSOCIATION RULE

Pruning in classification association rule mining is a crucial step that involves refining and simplifying the rule set to enhance predictive accuracy. By removing redundant or less informative rules, pruning ensures a more concise and interpretable model, reducing overfitting and improving generalization to unseen data[11]. This process is instrumental in crafting a predictive model that strikes a balance between complexity and performance, thereby optimizing the efficiency and reliability of association rule-based predictions.

RANDOM FOREST ALGORITHM

The Random Forest algorithm is a powerful tool in machine learning for prediction tasks. It operates by constructing multiple decision trees during training and outputs the mode of the classes for classification or the mean prediction for regression[13]. Known for its ability to handle diverse datasets and mitigate overfitting, Random Forest has proven effective in predicting outcomes across various domains, from healthcare diagnoses to financial forecasting.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 12, December 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1112022 |

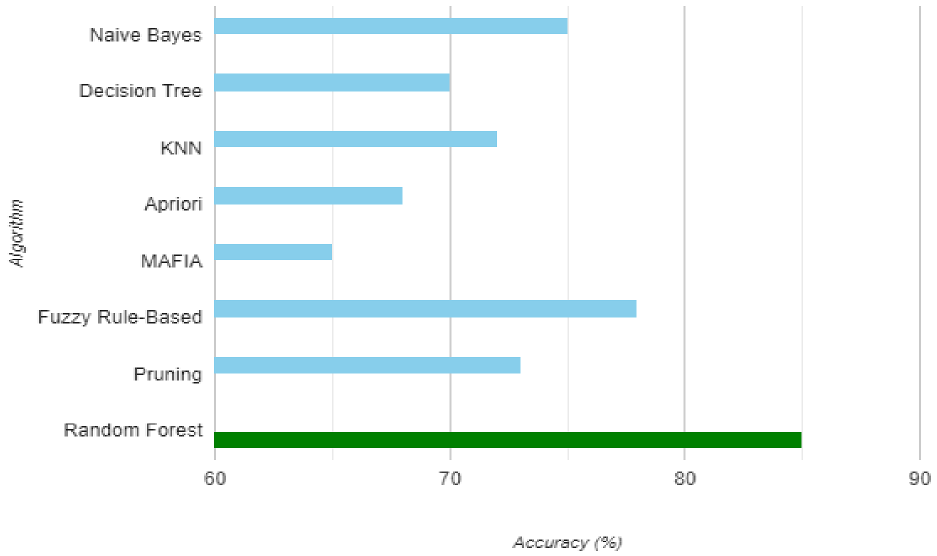


Figure .2 Comparison of Classification Algorithm

The above graph points that Random Forest Algorithm provides the highest accuracy for prediction.Hence ,Random forest Algorithm can be chosen for heart disease prediction.Random Forest is an ensemble learning algorithm used for both classification and regression tasks. It builds multiple decision trees during training and outputs the mode of the classes (for classification) or the mean prediction (for regression) of the individual trees. Each tree in the forest is constructed using a random subset of the data and a random subset of features, which enhances the model's robustness and reduces overfitting.

In the context of heart disease prediction, Random Forest offers distinct advantages over other algorithms. Its ability to handle a variety of features and complex interactions within the data makes it particularly well-suited for medical datasets with diverse variables. Moreover, Random Forest provides built-in mechanisms to assess feature importance, aiding in the identification of key factors contributing to heart disease, a crucial aspect in the field of predictive healthcare.

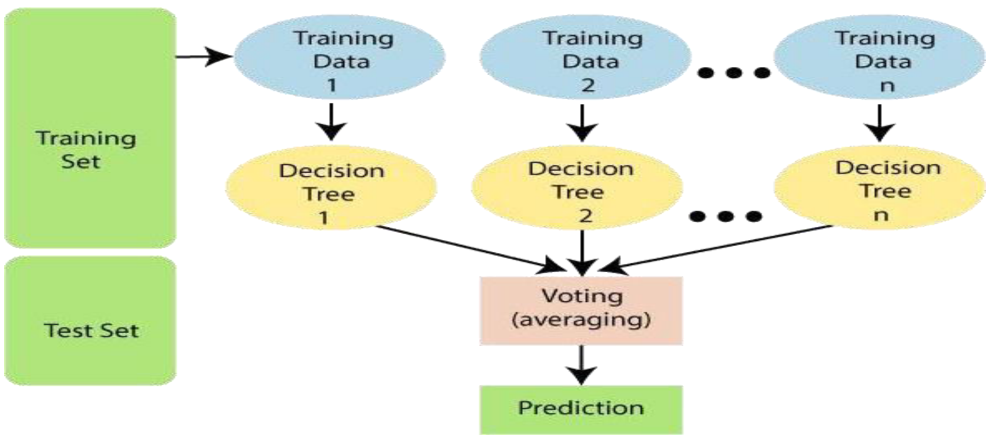


Figure .3 Random Forest Algorithm

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 12, December 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1112022 |

V. CONCLUSION

In conclusion, the escalating prevalence of heart disease in contemporary times has prompted extensive research into predictive methodologies. A thorough study of various related works has revealed the pressing need for accurate and early detection methods to curb the rising tide of cardiovascular issues. Among the myriad algorithms evaluated, the Random Forest algorithm has emerged as a standout performer, exhibiting superior accuracy in predicting heart disease. Its capacity to handle diverse datasets, consider complex interactions, and offer insights into feature importance positions it as a valuable tool for early detection. By leveraging the strengths of Random Forest, there is promising potential to revolutionize the landscape of heart disease prediction, facilitating proactive interventions and ultimately contributing to improved public health outcomes.

REFERENCES

[1] Jyoti Soni, Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction, International Journal of Computer Applications (0975 – 8887) Volume 17– No.8, March 2011.

[2] P.K. Anooj, Clinical decision support system: Risk level prediction of heartdisease using weighted fuzzy rules, Journal of King Saud University – Computer and Information Sciences (2012) 24, 27–40.

[3] Nidhi Bhatla and Kiran Jyoti, An Analysis of Heart Disease Prediction using Different Data Mining Techniques, International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 8, October - 2012 ISSN: 2278-0181.

[4] Aditya Methaila, Prince Kansal, Himanshu Arya, Pankaj Kumar, Early Heart Disease Prediction Using Data Mining Echniques,Sundarapandian et al. (Eds) : CCSEIT, DMDB, ICBB, MoWiN, AIAP – 2014 pp. 53–59, 2014. © CS & IT-CSCP 2014 DOI : 10.5121/csit.2014.4807.

[5] Shimpy Goyal and Dr. Rajender Singh Chhillar , A Literature Survey on Applications of Data Mining Techniques to Predict Heart Diseases,International Journal of Engineering Sciences Paradigms and Researches (IJESPR) (Vol. 20, Issue 01) and (Publishing Month: May 2015)

[6] 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 Vol. 2(10), 2010, 5370-5376

[7] M.Akhil jabbar, Dr.Priti Chandra, Dr.B.L Deekshatulu " Heart Disease Prediction System using Associative Classification and Genetic Algorithm", ICECIT, 2012.

[8] R. Sethukkarasi and Kannan, "An Intelegent System for Mining Temporal rules inClinical database using Fuzzy neural network", European Journal of Scientific Research, ISSN 1450-216, Vol 70(3), pp 386-395, 2012.

[9] Mohammed Abdul Khaleel, Sateesh Kumar Pradhan, " Finding Locally Frequent Diseases Using Modified Apriori Algorithm", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 10, October 2013.

[10] Boshra Bahrami, Mirsaeid Hosseini Shirvani, "Prediction and Diagnosis of HeartDisease by Data Mining Techniques", Journal of Multidisciplinary Engineering Science and Technology (JMEST) ISSN: 3159-0040 Vol. 2 Issue 2, February–2015.

[11] Deepika .N, "Association Rules for Classifiaction of Heart Attack Patients",IJAEST, Vol 11(2), pp 253-257, 2011.

[12] Cleveland database: http://archive.ics.uci.edu/ml/datasets/Heart+Disease.

[13] Heart Disease Prediction Using Random Forest Algorithm.Kompella Sri Charan1, Kolluru S S N S Mahendranath,K. Sri charan,4-177, Main Road, Amalapuram, Andhra Pradesh,K.S S N S Mahendranath, Nellore, Andhra Pradesh,

[14] Statlog database: http://archive.ics.uci.edu/ml/machine-learning databases/statlog/heart