

Heart Disease Identification Method using Machine Learning Classification in E-Healthcare

¹Mrs.R.Aarathi, ²S.Vijayasimman, ³P.Kathiravan, ⁴S.Saravanan, ⁵M.Velusamy

¹Assistant professor, ^{2,3,4,5}Students
Department of Computer Science and Engineering,
Nandha College of Technology, Perundurai, Erode, Tamilnadu, India

Abstract: Coronary artery heart Disease (CAD) is caused by atherosclerosis in coronary arteries and results in cardiac arrest and heart attack. For diagnosis of CAD, angiography is used which is a costly time consuming and highly technical invasive method. Researchers are therefore, prompted for alternative methods such as machine learning algorithms that could use non-invasive clinical data for the heart Disease diagnosis and assessing its severity. In this study, we present a novel hybrid method for CAD diagnosis, including risk factor identification using correlation based feature subset (CFS) selection with particle swarm optimization (PSO) search method and K-Means clustering algorithms. Supervised learning algorithms such as multi-layer perceptron (MLP), multinomial logistic regression (MLR), fuzzy unordered rule induction algorithm (FURIA) and C4.5 are then used to model CAD cases. We tested this approach on clinical data consisting of 26 features and 335 instances collected at the Department of Cardiology, Indira Gandhi Medical College, and Shimla, India. MLR achieves highest prediction accuracy of 88.4 %. We tested this approach on benchmarked Cleveland heart Disease data as well. In this case also, MLR, outperforms other techniques. Proposed hybridized model improves the accuracy of classification algorithms from 8.3 % to 11.4 % for the Cleveland data. The proposed method is, therefore, a promising tool for identification of CAD improved prediction accuracy.

1. Introduction

Data Mining is an analytic process designed to explore data (usually large amounts of data - typically business or market related) in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. Machine learning (ML) that improve naturally through experience. It is viewed as a subset of man-made brainpower. AI calculations construct a model dependent on example information, known as "preparing information", to settle on expectations or choices without being unequivocally customized to do as such. Illnesses under the coronary illness umbrella incorporate vein sicknesses, for example, coronary conduit infection; heart musicality issues (arrhythmias); and heart abandons you're brought into the world with (inherent heart deserts), among others. The expression "coronary illness" is regularly utilized reciprocally with the expression "cardiovascular infection." Prescient demonstrating utilizes measurements to foresee outcomes.[1] Most frequently the occasion one needs to anticipate is later on, yet prescient displaying can be applied to an obscure occasion, paying little mind to happened.

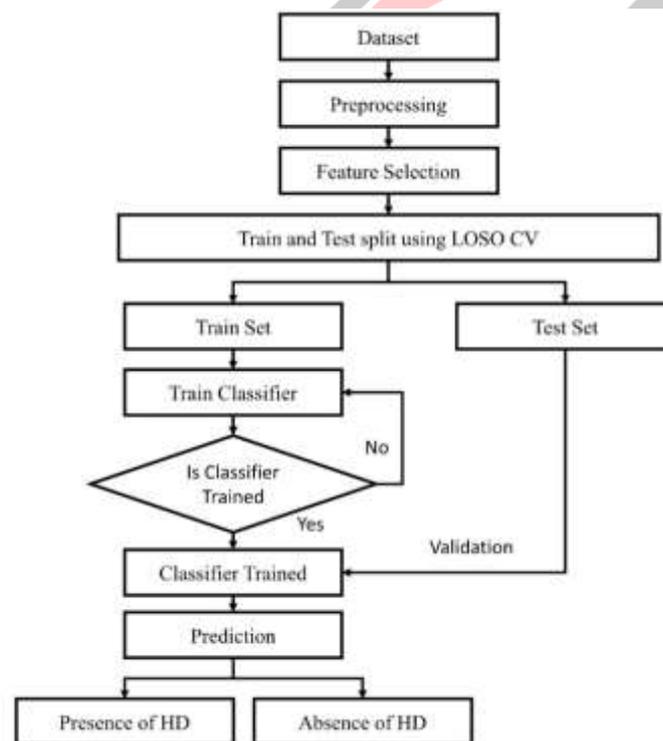


Figure 1 Architecture of heart disease system

2. Literature Review

In literature various machine learning based diagnosis techniques have been proposed by researchers to diagnosis HD. This research study present some existing machine learning based diagnosis techniques in order to explain the important of the proposed work. Detrano *et al.* [11] developed HD classification system by using machine learning classification techniques and the performance of the system was 77% in terms of accuracy. Cleveland dataset was utilized with the method of global evolutionary and with features selection method. In another study Gudadhe *et al.* developed a diagnosis system using multi-layer Perceptron and support vector machine (SVM) algorithms for HD classification and achieved accuracy 80.41%. Humar *et al.* designed HD classification system by utilizing a neural network with the integration of Fuzzy logic. The classification system achieved 87.4% accuracy. Resul *et al.* developed an ANN ensemble based diagnosis system for HD along with statistical measuring system enterprise miner (5.2) and obtained the accuracy of 89.01%, sensitivity 80.09%, and specificity 95.91%. Akil *et al.* designed a ML based HD diagnosis system. ANN-DBP algorithm along with FS algorithm and performance was good. Palaniappan *et al.* [7] proposed an expert medical diagnosis system for HD identification. In development of the system the predictive model of machine learning, such as navies bays (NB), Decision Tree (DT), and Artificial Neural Network were used. The 86.12% accuracy was achieved by NB, ANN accuracy 88.12% and DT classifier achieved 80.4% accuracy. Olaniyi *et al.* developed a three-phase technique based on the artificial neural network technique for HD prediction in angina and achieved 88.89% accuracy. Samuel *et al.* developed an integrated medical decision support system based on artificial neural network and Fuzzy AHP for diagnosis of HD. The performance of the proposed method in terms of accuracy was 91.10%. Liu *et al.* proposed a HD classification system using relief and rough set techniques. The proposed method achieved 92.32% classification accuracy. In proposed a HD identification method using feature selection and classification algorithms. Sequential Backward Selection Algorithm (SBS FS) for Features Selection. The classifier K-Nearest Neighbor (K-NN) performance has been checked on full and on selected features set. The proposed method obtained high accuracy. In another study MOHAN *et al.* Designed a HD prediction method by using hybrid machine learning techniques.

3. Existing System

In Existing medical systems, including hospital management systems and decision making systems, focus on collecting and mining the entire medical data. The entire patient records are loaded and all factors are considered. The C4.5 model that we previously developed. On the basis of this new measure, a data mining algorithm was developed to mine the causal relationship between drugs and their associated with risk for CAD heart Disease.

4. Drawbacks

- Existing systems have failed to utilize and understand the importance of misdiagnosis.
- A very important attribute which interconnects and addresses all these issues.
- It varies from patient's medical history, climatic conditions, neighborhood, and various other factors.

5. Proposed system

In the proposed work user will search for the heart Disease diagnosis (heart Disease and treatment related information) by giving symptoms as a query in the search engine. These symptoms are preprocessed to make the further process easier to find the symptoms keyword which helps to identify the heart Disease quickly. CFS+PSO are a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. This feature has been identified as the most suitable for the present system.

6. Advantage

- It is easy to extract signatures from individual data instances, as their structures. Just collect the symptoms that enough to scaling samples.
- Can easily predict the heart Disease level and severity easily using range level of queries.
- The probability of vocabulary gap between diverse health seekers makes the data more consistent compared to other formats of health data.

7. Materials and Method

7.1 Data Visualization and Pre-Processing

The Wisconsin Prognostic Cleve Land Train Dataset is downloaded from the UCI Machine Learning Repository website and saved as a text file. This file is then imported into Excel spreadsheet and the values are saved with the corresponding attributes as column headers. The missing values are replaced with appropriate values.

7.2 Dimensionality Reduction (CFS+PSO)

Feature selection, by identifying the most salient features for learning, focuses a learning algorithm on those aspects of the data most useful for analysis and future prediction. Feature selection is a process of selecting a subset of relevant features from a large number of original features to achieve similar or better classification performance and improve the computation efficiency

7.3 Model for Cad Identification

Multinomial logistic regression model (MLR): Multinomial logistic regression is known by a variety of other names, including polytomous LR, multiclass LR, softmax regression, multinomial logit, maximum entropy (MaxEnt) classifier, and conditional maximum entropy model.

7.4 Risk Prediction

It provides reproducible and objective diagnosis, and hence can be a valuable adjunct tool in clinical practices. Then the data are clustered using MLP, MLG, FURIA and C4.5 using all the features of CAD data. Risk forecast instruments are created to recognize

patients in danger and to encourage doctor dynamic.

7.5 Classification Algorithms

Then the data are clustered using MLP, MLG, FURIA and C4.5 using all the features of CAD. In bunching, the thought isn't to foresee the objective class as in order, it's additionally attempting to assemble the comparative sort of things by thinking about the most fulfilled condition, all the things in a similar gathering should be comparable and ought no two distinctive gathering things to not be comparative. Experiment results demonstrate the superiority of the proposed method with regard to prediction accuracy of CAD with the features selected by CFS & PSO, we need only a few clinical data to apply this model.

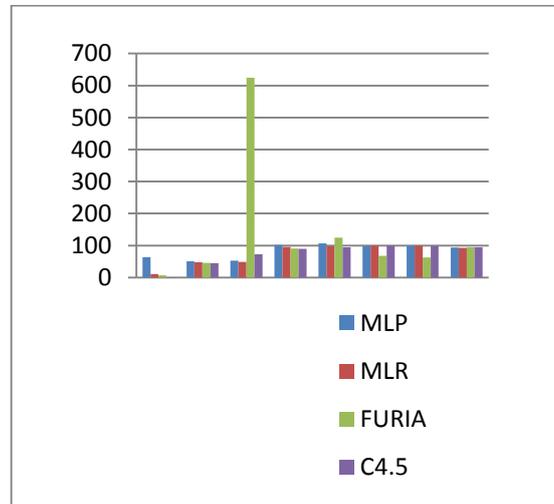


Figure 2 Performance of MLP, MLG, FURIA and C4.5 using all the features of CAD data.

8. Conclusion

Clinical finding is a significant region of exploration which assists with recognizing the event of a coronary illness. The framework, utilizing different methods referenced, will thus uncover the root coronary illness alongside the arrangement of most plausible heart Diseases which have comparative side effects. The information base utilized is a portrayal data set so to decrease the dataset tokenization, separating and stemming is finished. The venture presents a novel mixture model to recognize and affirm CAD cases requiring little to no effort by utilizing clinical information that can be effectively gathered at clinics. Intricacy of framework is diminished by decreasing the dimensionality of the informational collection with PSO. It gives reproducible and target finding, and subsequently can be a significant extra device in clinical practices. Results are equivalently, encouraging and along these lines the proposed half and half technique will be useful in coronary illness diagnostics. Trial results exhibit the predominance of the proposed half breed technique concerning forecast precision of CAD.

References

- [1] S. I. Ansarullah and P. Kumar, "A systematic literature review on cardiovascular disorder identification using knowledge mining and machine learning method," *Int. J. Recent Technol. Eng.*, vol. 7, no. 6S, pp. 1009–1015, 2019
- [2] A. U. Haq, J. P. Li, J. Khan, M. H. Memon, S. Nazir, S. Ahmad, G. A. Khan, and A. Ali, "Intelligent machine learning approach for effective recognition of diabetes in E-healthcare using clinical data," *Sensors*, vol. 20, no. 9, p. 2649, May 2020
- [3] A. U. Haq, J. Li, M. H. Memon, M. H. Memon, J. Khan, and S. M. Marium, "Heart disease prediction system using model of machine learning and sequential backward selection algorithm for features selection," in *Proc. IEEE 5th Int. Conf. Converg. Technol. (ICT)*, Mar. 2019, pp. 1–4
- [4] U. Haq, J. Li, M. H. Memon, J. Khan, and S. U. Din, "A novel integrated diagnosis method for breast cancer detection," *J. Intell. Fuzzy Syst.*, vol. 38, no. 2, pp. 2383–2398, 2020.
- [5] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019.
- [6] G. G. N. Geweid and M. A. Abdallah, "A new automatic identification method of heart failure using improved support vector machine based on duality optimization technique," *IEEE Access*, vol. 7, pp. 149595–149611, 2019.
- [7] A. Ul Haq, J. Li, Z. Ali, J. Khan, M. H. Memon, M. Abbas, and S. Nazir, "Recognition of the Parkinson's disease using a hybrid feature selection approach," *J. Intell. Fuzzy Syst.*, vol. 39, pp. 1–21, May 2020, doi: 10.3233/JIFS-200075
- [8] R. Sivaranjani, V. S. Naresh, and N. V. Murthy, "4 coronary heart disease prediction using genetic algorithm based decision tree," *Intell. Decis. Support Syst., Appl. Signal Process.*, vol. 4, p. 71, Oct. 2019
- [9] A. U. Haq, J. Li, J. Khan, M. H. Memon, S. Parveen, M. F. Raji, W. Akbar, T. Ahmad, S. Ullah, L. Shoista, and H. N. Monday, "Identifying the predictive capability of machine learning classifiers for designing heart disease detection system," in *Proc. 16th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process.*, Dec. 2019, pp. 130–138.
- [10] A. U. Haq, J. P. Li, M. H. Memon, J. Khan, A. Malik, T. Ahmad, A. Ali, S. Nazir, I. Ahad, and M. Shahid, "Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson's disease using voice recordings," *IEEE Access*, vol. 7, pp. 37718–37734, 2019.