Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue 2, 2024

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

[1] B.N.S GUPTA, [2] YELURI ASHISH

[1]Associate Professor, S.V.K.P & Dr K.S. Raju Arts & Science College (A), PENUGONDA, W.G.District, Andhra Pradesh, bnsgupta@gmail.com

[2]PG, scholar, S.V.K.P & Dr K.S. Raju Arts & Science College (A), Penugonda,W.G.District,Andhra Pradesh, yeluriashish98@gmail.com

ABSTRACT

Heart disease is one of the complex diseases and globally many people suffered from this disease. On time and efficient identification of heart disease plays a key role in healthcare, particularly in the field of cardiology. In this article, we proposed an efficient and accurate system to diagnosis heart disease and the system is based on machine learning techniques. The system is developed based on classification algorithms includes Support vector machine, Logistic regression, Artificial neural network, K-nearest neighbor, Naïve bays, and Decision tree while standard features selection algorithms have been used such as Relief, Minimal redundancy maximal relevance, Least absolute shrinkage selection operator and Local learning for removing irrelevant and redundant features. We also proposed novel fast conditional mutual information feature selection algorithm to solve feature selection problem. The features selection algorithms are used for features selection to increase the classification accuracy and reduce the execution time of classification system. Furthermore, the leave one subject out cross-validation method has been used for learning the best practices of model assessment and for hyper parameter tuning. The performance measuring metrics are used for assessment of the performances of the classifiers. The performances of the classifiers have been checked on the selected features as selected by features selection algorithms. The experimental results show that the proposed feature selection algorithm (FCMIM) is feasible with classifier support

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue 2, 2024

vector machine for designing a high-level intelligent system to identify heart disease

1.INTRODUCTION

Heart disease (HD) is the critical health issue and numerous people have been suffered by this disease around the world [1]. The HD occurs with common symptoms of breath shortness, physical body weakness and, feet are swollen [2]. Researchers try to come across an efficient technique for the detection of heart disease, as the current diagnosis techniques of heart disease are not much effective in early time identification due to several reasons, such as accuracy and execution time [3]. The diagnosis and treatment of heart disease is extremely difcult when modern technology and medical experts are not available [4]. The effective diagnosis and proper treatment can save the lives of many people [5]. According to the European Society of Cardiology, 26 million approximately people of HD were diagnosed and diagnosed 3.6 million annually [6]. Most of the people in the United States are suffering from heart disease [7]. Diagnosis of HD is traditionally done by the analysis of the medical history of the patient, physical examination report and analysis of concerned symptoms by a physician. But the results obtained from this diagnosis method are not accurate in identifying the patient of HD. Moreover, it is expensive and computationally difcult to analyze [8].

Thus, to develop a noninvasive diagnosis system based on classiers of machine learning (ML) to resolve these issues. Expert decision system based on machine learning classiers and the application of artificial fuzzy logic is effectively diagnosis the HD as a heart disease data set was used by various researchers [11] and [12] for the identification problem of HD. The machine learning predictive models need proper data for training and testing. The performance of machine learning model can be increased if balanced dataset is use for training and testing of the model. Furthermore, the model predictive capabilities can improved by using proper and related features from the data. Therefore, data balancing and feature selection is significantly important for model performance improvement. In literature various diagnosis techniques have been proposed by various researchers, however these techniques are not effectively diagnosis HD. In order to improve the predictive

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue 2, 2024

capability of machine learning model data preprocessing is important for data standardization. Various Preprocessing techniques such removal of missing feature value instances from the dataset, Standard Scalar (SS), Min-Max Scalar etc.

The feature extraction and selection techniques are also improve model performance. Various feature selection techniques are mostly used for important feature selection such as, Least-absolute-shrinkage-selection-operator (LASSO), Relief, Minimal-Redundancy-Maximal-Relevance (MRMR), Local-learning-based features-selection (LLBFS), Principle component Analysis (PCA), Greedy Algorithm (GA), and optimization methods, such as Anty Conley Optimization (ACO), fruit y optimization (FFO), Bacterial Foraging Optimization (BFO) etc. Similarly Yun et al. [13] presented different techniques for different type of feature selection, such as feature selection for high-dimensional small sample size data, large-scale data, and secure feature selection. They also discussed some important topics for feature selection have emerged, such as stable feature selection, multi view feature selection, distributed feature selection, multi-label feature selection, online feature selection, and adversarial feature selection. Jundong et al. [14] discussed the challenges of feature selection (FS) for big data. It is necessary to decrease the dimensionality of data for various learning tasks due to the curse of dimensionality.

2.LITERATURE SURVEY

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. 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.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue 2, 2024

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. 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

3. EXISTING SYSTEM

- 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.* [22] 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.* [23] 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.* [19] 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.* [24] designed a ML based HD diagnosis system. ANN-DBP algorithm along with FS algorithm and performance was good.
- Palaniappan *et al.* [17] 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.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue 2, 2024

Disadvantages

In the existing work, the system is poor performance in which the low performance LOCAL LEARNING BASED FEATURES SELECTION ALGORITHM is used. This system is less performance due to Lack of Heart disease classification Techniques

4. PROPOSED SYSTEM.

The system also proposed fast conditional mutual information (FCMIM) features selection algorithm for features selection. Leave-one-subject-out cross-validation (LOSO) technique has been applied to select the best hyper-parameters for best model selection. Apart from this, different performance assessment metrics have been used for classifiers performances evaluation. The proposed method has been tested on Cleveland HD dataset. Furthermore, the performance of the proposed technique have been compared with state of the art existing methods in the literature, such as NB [17], Three phase ANN (Artificial neural Network) diagnosis system [18], Neural network ensembles (NNE) [19], ANN-Fuzzy-AHP diagnosis system (AFP) [20], Adaptive-weighted-Fuzzy-system-ensemble (AWFSE) [21]. The research study has the following contributions.

Firstly, the authors try to address the problem of features selection by employing pre-processing techniques and standard state of the art four features selection algorithms such as Relief, mRMR, LASSO, and LLBFS for appropriate subset of features and then applied these features for effective training and testing of the classifiers that identify which feature selection algorithm and classifier gives good results in term of accuracy and computation time.

Secondly, the authors proposed fast conditional mutual information (FCMIM) FS algorithm for feature selection and then these features are input to classifiers for improving prediction accuracy and reducing computation time. The classifiers performances have been compared on features selected by the standard state of the art FS algorithms with the selected features of the proposed FS algorithm.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue 2, 2024

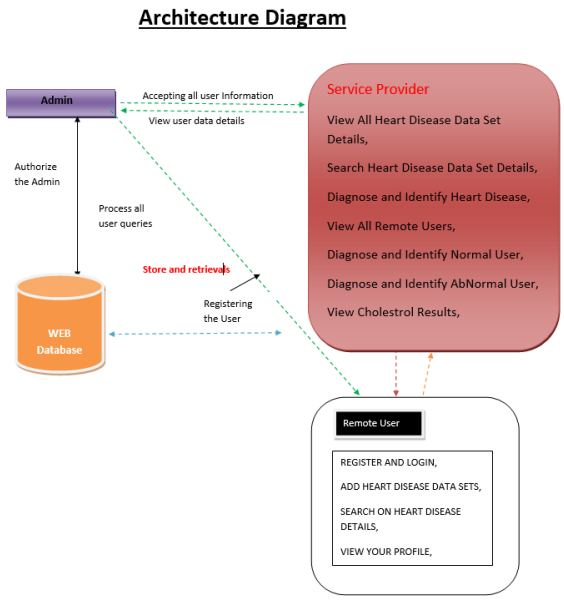
Thirdly, identify weak features from the dataset which affect the performance of the classifiers.

Finally, suggests that heart disease identification system (FCMIM-SVM) effectively identify the HD.

Advantages

The system is fast and reliable due to presence of feature selection algorithm (FCMIM). The system is more effective due to TANDARD STATE OF THE ART FEATURES SELECTION ALGORITHM

5. SYSTEM ARCHITECTURE



6. MODULES

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as

View All Heart Disease Data Set Details,Search Heart Disease Data Set Details,Diagnose and Identify Heart Disease,View All Remote Users,Diagnose and Identify Normal User,Diagnose and Identify AbNormal User,View Cholestrol Results,View Heart Beat Results.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login

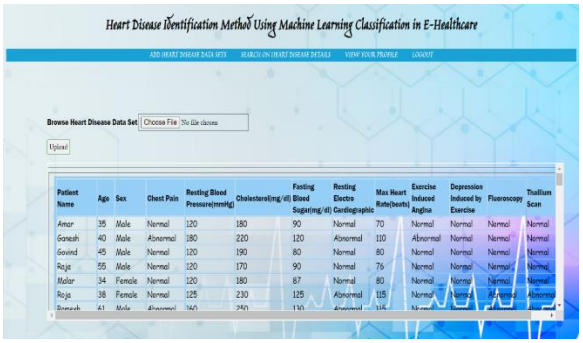
International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue 2, 2024

by using authorized user name and password. Once Login is successful user will do some operations like ADD HEART DISEASE DATA SETS, SEARCH ON HEART DISEASE DETAILS, VIEW YOUR PROFILE.

7. SCREENS



8. CONCLUSION

In this study, an efficient machine learning based diagnosis system has been developed for the diagnosis of heart disease. Machine learning classiers include LR, K-NN, ANN, SVM, NB, and DT are used in the designing of the system. Four standard feature selection algorithms including Relief, MRMR, LASSO, LLBFS, and proposed a novel feature selection algorithm FCMIM used to solve feature selection problem. LOSO cross-validation method is used in the system for the best hyper parameters

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue 2, 2024

selection. The system is tested on Cleveland heart disease dataset. Furthermore, performance evaluation metrics are used to check the performance of the identification system. According to Table 15 the specicity of ANN classier is best on Relief FS algorithm as compared to the specicity of MRMR, LASSO, LLBFS, and FCMIM feature selection algorithms. Therefore for ANN with relief is the best predictive system for detection of healthy people. The sensitivity of classier NB on selected features set by LASSO FS algorithm also gives the best result as compared to the sensitivity values of Relief FS algorithm with classier SVM (linear). The classier Logistic Regression MCCis 91% on selected features selected by FCMIM FS algorithm. The processing time of Logistic Regression with Relief, LASSO, FCMIM and LLBFS FS algorithm best as compared to MRMR FS algorithms, and others classiers. Thus the experimental results show that the proposed features selection algorithm select features that are more effective and obtains high classification accuracy than the standard feature selection algorithms. According to feature selection algorithms, the most important and suitable features are Thallium Scan type chest pain and Exercise-induced Angina. All FS algorithms results show that the feature Fasting blood sugar (FBS) is not a suitable heart.

9. REFERENCE

[1] A. L. Bui, T. B. Horwich, and G. C. Fonarow, ``Epidemiology and risk prole of heart failure," Nature Rev. Cardiol., vol. 8, no. 1, p. 30, 2011.

[2] M. Durairaj and N. Ramasamy, ``A comparison of the perceptive approaches for preprocessing the data set for predicting fertility success rate," Int. J. Control Theory Appl., vol. 9, no. 27, pp. 255260, 2016.

[3] L. A. Allen, L.W. Stevenson, K. L. Grady, N. E. Goldstein, D. D. Matlock, R. M. Arnold, N. R. Cook, G. M. Felker, G. S. Francis, P. J. Hauptman, E. P. Havranek, H. M. Krumholz, D. Mancini, B. Riegel, and J. A. Spertus, ``Decision making in advanced heart failure: A scientic statement from the American heart association," Circulation, vol. 125, no. 15, pp. 19281952, 2012.

[4] S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, ``Innovative articial

ISSN 2319-5991 www.ijerst.com
Vol. 17, Issue 2, 2024

neural networks-based decision support system for heart diseases diagnosis,"
J. Intell. Learn. Syst. Appl., vol. 5, no. 3, 2013, Art. no. 35396.
[5] Q. K. Al-Shayea, ``Articial neural networks in medical diagnosis," Int. J.
Comput. Sci. Issues, vol. 8, no. 2, pp. 150154, 2011.
[6] J. Lopez-Sendon, ``The heart failure epidemic," Medicographia, vol. 33,
no. 4, pp. 363369, 2011.
[7] P. A. Heidenreich, J. G. Trogdon, O. A. Khavjou, J. Butler, K. Dracup,
M. D. Ezekowitz, E. A. Finkelstein, Y. Hong, S. C. Johnston, A. Khera,
D. M. Lloyd-Jones, S. A. Nelson, G. Nichol, D. Orenstein, P.W. F.Wilson,
and Y. J. Woo, ``Forecasting the future of cardiovascular disease in the
united states: A policy statement from the American heart association,"
Circulation, vol. 123, no. 8, pp. 933944, 2011.
[8] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, ``Nonlinear
speech analysis algorithms mapped to a standard metric achieve clinically
useful quantication of average Parkinson's disease symptom severity,"
J. Roy. Soc. Interface, vol. 8, no. 59, pp. 842855, 2011. 9] S. I. Ansarullah and P. Kumar, ``A systematic literature review on cardiovascular disorder identication using knowledge mining and
machine learning method," Int. J. Recent Technol. Eng., vol. 7, no. 6S,
pp. 10091015, 2019.
[10] S. Nazir, S. Shahzad, S. Mahfooz, and M. Nazir, ``Fuzzy logic based
decision support system for component security evaluation," Int. Arab J.
Inf. Technol., vol. 15, no. 2, pp. 224231, 2018.
[11] R. Detrano, A. Janosi,W. Steinbrunn, M. Psterer, J.-J. Schmid, S. Sandhu,
K. H. Guppy, S. Lee, and V. Froelicher, ``International application of a new
probability algorithm for the diagnosis of coronary artery disease," Amer.
J. Cardiol., vol. 64, no. 5, pp. 304310, Aug. 1989.
[12] J. H. Gennari, P. Langley, and D. Fisher, ``Models of incremental concept
formation," Artif. Intell., vol. 40, nos. 13, pp. 1161, Sep. 1989.
[13] Y. Li, T. Li, and H. Liu, ``Recent advances in feature selection and its
applications," Knowl. Inf. Syst., vol. 53, no. 3, pp. 551577, Dec. 2017.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue 2, 2024

[14] J. Li and H. Liu, ``Challenges of feature selection for big data analytics,"
IEEE Intell. Syst., vol. 32, no. 2, pp. 915, Mar. 2017.

[15] L. Zhu, J. Shen, L. Xie, and Z. Cheng, ``Unsupervised topic hypergraph
hashing for efcient mobile image retrieval," IEEE Trans. Cybern., vol. 47,
no. 11, pp. 39413954, Nov. 2017.