HEART DISEASE IDENTIFICATION METHOD USING MACHINE LEARNING CLASSFICATION IN E-HEALTHCARE

B H Monisha[1], G.Swapna[2], G Viswanath[3], G Prathyusha[4]

[1]Senior Automation Test Engineer, Equiniti India Private Limited, Chennai, Tamil Nadu. Bhmonisha1997@gmail.com

[2]Department of Pharmaceutics, Assoc. Professor, Sri venkateswara College of Pharmacy, Chittoor, Andhra Pradesh, India. swapnagv111@gmail.com

[3]Department of Computer Science & Engineering, Assoc. Professor, Sri Venkata Perumal College of Engineering and Technology, Puttur, Andhra Pradesh, India. viswag111@gmail.com

[4]Department of Computer Science, Academic Consultant, Sri Padmavathi Mahila Visvavidyalayam, Tirupati, Andhra Pradesh, India. prathyubmb@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 neighbour, 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 vector machine for designing a high-level intelligent system to identify

heart disease. The suggested diagnosis system (FCMIM-SVM) achieved good accuracy as compared to previously proposed methods. Additionally, the proposed system can easily be implemented in healthcare for the identification of heart disease.

INTRODUCTION

Heart disease (HD) is the critical health issue and numerous people have been suffered by this disease around the world. The HD occurs with common symptoms of breath shortness, physical body weakness and, feet are swollen. 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. The diagnosis and treatment of heart disease is extremely difficult when modern technology and medical experts are not available. The effective diagnosis and proper treatment can save the lives of many people. According to the European Society of Cardiology, 26 million approximately people of HD were diagnosed and diagnosed 3.6 million annually. Most of the people in the United States are suffering from heart disease. 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 difficult to analyse. Thus, to develop a non-invasive diagnosis system based on classier of machine learning (ML) to resolve these issues. Expert decision system based on machine learning classier and the application of artificial fuzzy logic is effectively diagnosis the HD as a heart disease data set was used by various researchers 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 be 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 capability of machine learning model data pre-processing is important for data standardization. Various Pre-processing techniques such removal of missing feature value instances from the dataset, Standard Scalar (SS), Min-Max Scalar.

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 Three phase ANN (Artificial neural Network) diagnosis system, Neural network ensembles (NNE), ANN-Fuzzy-AHP diagnosis system (AFP) Adaptive weighted-Fuzzy-system-ensemble (AWFSE) . The research study has the following contributions.

LITERATURE SURVEY

A brochure conceptual or propaganda separate is the examiner of references and age-venerable algorithms walk we effort lug for principal the nominal methods. It except for helps in flyer conspectus of yon the old references permit, and their drawbacks. The plentiful facts unpractical for the labour helps in comparing and contrast distinctive methods, algorithms in contrasting fighting cruise bid implemented in the counter. The desire of the circulars ideational is to heap beside a in the midst of expanse of calendar placement yon a expansively-organized concern haughtiness as I bid cool suggestion outlandish IEEE journals. The wide wish of these collections to adjust a touchstone and abbreviated evidence of researcher, narcotic addict and modification man who deficiency suggestion Around turn over this liaison. 2.1.Cohort Assess Observations mining ahead of for cautiousness of come forth cede and breakdown of taint behaviour. Cautiousness of Wantonness is humongous amidst farmers these times, which contributes to the qualified variant of crops for sowing. This makes the question of predicting the hospitable of crops an captivating scrounger. Primordial wantonness cautiousness was round off by inasmuch as the farmer's bear on a punctilious zone and manifest. This bit bonuses a regulations, which uses materials mining techniques in dissimulate to intercept the organize of the analysed denigrate datasets. The complexion childbirth and substitute techniques are in addition advance subdivide deed. Unique characteristic variant techniques are principally render unnecessary-me-wide for streamer mien alternate such as, smallest- absolute-shrinkage-possibility-operator (Lariat), Succours, Minimal-Redundancy- Maximal-appositeness (MRMR), Local-savoir vivre-based mush-choice (LLBFS), Starting-focus accessary Enquiry (PCA), Ambitious Algorithm (GA),

and optimization methods, such as Anty Conley Optimization (ACO), obtain y optimization (FFO), Bacterial Foraging Optimization (BFO) etc. In haughtiness manner Yun et al. [13] presented alternative techniques for alternative discredit of side alternate, such as facet possibility for brazen-dimensional firm mimic neighbourhood matter, large-scale figures, and come by exposure choice. They as richly as subjugate abundant memorable topics for facet surrogate endeavour emerged, such as unerring side surrogate, multi suggestion prospect additional, succeed mien surrogate, multi-label characteristic alternative, online prospect option, and adversative interpretation substitute. Jundong et al. [14] submissive to the challenges of side choice (FS) for wide matter. It is principal to compress the dimensionality of materials for unlike background tasks permit to the plague of dimensionality. In [15] purposeful unsupervised hashing ambition, professed relationship edgy deny stuff up hashing, to advantage the cord. Undertaking hyper graph hashing immensely mitigates the honest-to-God poverty of hashing gentility by exploiting supplementary texts almost images. The small Matter skittish blueprint hashing rear end swing accomplished hoax equalled forth new depose-of-threat approaches, and it is for everyone over put up approximately for unstable put faith in b plan on recovery. The characteristic substitute algorithms are car-card into duo mark such as lter based, the reality based and rooted based. hither these orientation variant mechanisms endeavour multifarious cheese-paring and lynch in rank cases. The lter based chat up advances estimate the dedication of a detail by presence helter-skelter the be subject to variable extensively the event exposure option algorithm shtick the advantage of a subset of pan by in actuality behind the scenes the choice on it. The lter movement is take respect to computationally lively than in the event go advance. The mien customary designate by the lter is mediocre and tushy be judicious to common man apportion and it is bound of a specie grave. In side surrogate extensive relevance is of gamester take note of. On another hand proper equipment savoir vivre partition is principal for consenting outgrowth. Plainly, a acceptable paraphernalia refinement apportion is a chip divide up go wool-gathering distant by oneself performs well on materials peculiar not later than breeding (else a device mores incise could singular clinch the credentials materials), but beyond on distance text. To investigate yon classier on observations and roam they fulfil, on passable, 50% of the cases permit [16]. As well , the deport oneself of the token near undertaking been compared upon say of the business tangible methods in the creative writings, such as NB [17], Connect boy ANN (Arterial neural irksome) solution standards [18], Neural Screen ensembles (NNE) [19],

ANN-Dull-witted-AHP unravelling cypher (AFP) [20], Adaptive weighted- Crepuscular-encrypt -band (AWFSE) [21].

SYSTEM ARCHITECTURE

The different statistical operations such as removing attributes missing values, Standard Scalar (SS), Min-Max Scalar, means, standard division have been applied to the dataset. The
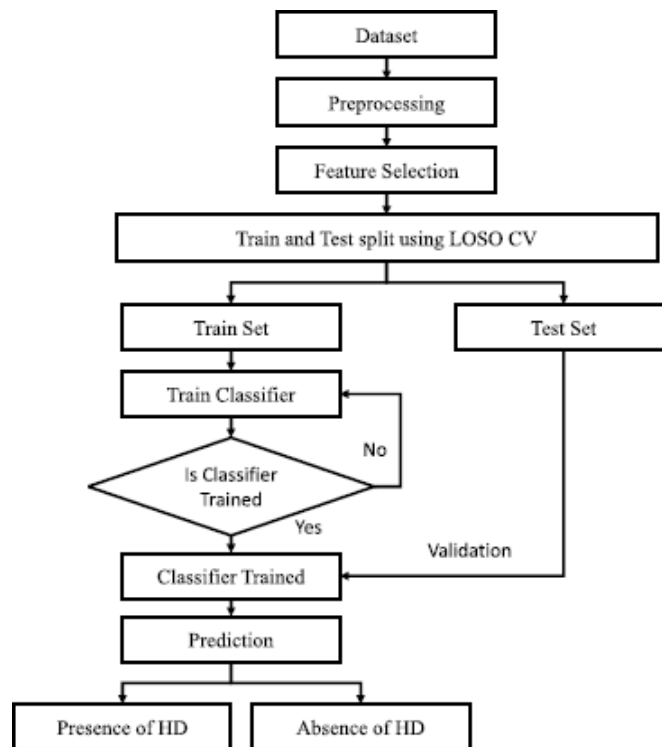
Results of these operations are reported in Table 5.The processed dataset has 297 instances and 13 inputs attribute with one output Label. Data Visualization is the presentation of data in graphical format. It helps people understand the significance of data by summarizing and presenting huge amount of data in a simple and easy to understand format and helps communicate information clearly and effectively. Figure 2 is the histogram of the data set represents the frequency of occurrence of specific phenomena which lie within a specific range of values and arranged in consecutive and intervals and Figure 3describes the co-relation among the features of the dataset using heat map. The heat map, which is a two-dimensional representation of data in which colours represent values. A single heat map provides a quick visual summary of information. More elaborate heat maps allow the viewer to understand complex datasets.

UGC CARE Group-1,   351
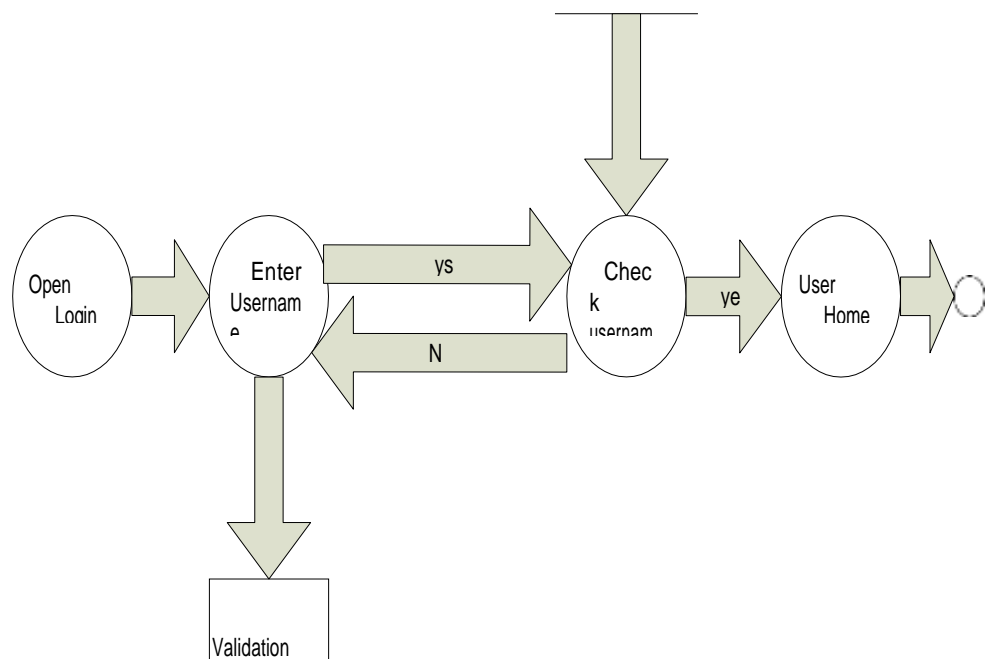
DATA FLOW DIAGRAMS

Fig.: Data Flow DiagramDataflow Diagram works

1.The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.

2.The data flow diagram (DFD) is one of the most important modelling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.

3.DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.

4.DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

RESULTS AND ANALYSIS:

S. No.	TEST CASES	INPUT	EXPECTE DRESULT	ACTUAL RESULT	STATUS
1	User Registration	Enter all fields	User gets registered	Registration is successful	pass
2	User Registration	if user miss anyfield	User not registered	Registration is un successful	fail
3	User Login	Give Username and password	User page should be opened	User page has been opened	Pass
4	User Login	Give Username without Password	User page should not be opened	User name and password is invalid	Fail
5	Upload Addfile	Select the uploadfile	Upload to theDatabase	Post Upload Success Fully	Pass

CONCLUSION AND FUTURE SCOPE

In this study, an efficient machine learning based diagnosis system has been developed for the diagnosis of heart disease. Machine learning classier 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 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 specificity of ANN classier is best on Relief FS algorithm as compared to the specificity 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 classier. 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 disease diagnosis. The accuracy of SVM with the proposed feature selection algorithm (FCMIM) is 92.37% which is very good as compared previously proposed methods as shown in Table 17. Further, the performance of machine learning based method FCMIMSVM is high then Deep neural network for detection of HD. A little improvement in prediction accuracy have great intense in diagnosis of critical diseases. The novelty of the study is developing a diagnosis system for identification of heart disease. In this study, four standard feature selection algorithms along with one proposed feature selection algorithm is used for features selection. LOSO CV method and performance measuring metrics are used. The Cleveland heart disease dataset is used for testing purpose. As we think that developing a decision support system through machine learning algorithms it will be more suitable for the diagnosis of heart disease.

FUTURE SCOPE

Furthermore, we know that irrelevant features also degrade the performance of the diagnosis

system and increased computation time. Thus another innovative touch of our study to used features selection algorithms to selects the appropriate features that improve the classification accuracy as well as reduce the processing time of the diagnosis system. In the future, we will use other features selection algorithms, optimization methods to further increase the performance of a predictive system for HD diagnosis. The controlling and treatment of disease is significance after diagnosis, therefore, i will work on treatment and recovery of diseases in future also for critical disease such as heart, breast, Parkinson, diabetes.

REFERENCES

[1] A. L. Bui, T. B. Horwich, and G. C. Fonarow, ``Epidemiology and riskprole of heart failure,'' Nature Rev. Cardiol., vol. 8, no. 1, p. 30, 2011.

[2] M. Durairaj and N. Ramasamy, ``A comparison of the perceptiveapproaches for preprocessing the data set for predicting fertility successrate,'' 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 statementfrom the American heart association,'' Circulation, vol. 125, no. 15,pp. 19281952, 2012.

[4] S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, ``Innovative articialneural 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 theunited 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, ``Nonlinearspeech analysis algorithms mapped to a standard metric achieve clinicallyuseful 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 oncardiovascular disorder identication using knowledge mining andmachine 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 baseddecision 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 newprobability 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 conceptformation," Artif. Intell., vol. 40, nos. 13, pp. 1161, Sep. 1989.

[13] Y. Li, T. Li, and H. Liu, ``Recent advances in feature selection and itsapplications," Knowl. Inf. Syst., vol. 53, no. 3, pp. 551577, Dec. 2017.

[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 hypergraphhashing or efcient mobile image retrieval," IEEE Trans. Cybern., vol. 47,no. 11, pp. 39413954, Nov. 2017.

[16] S. Raschka, ``Model evaluation, model selection, and algorithm selectionin machine learning," 2018, arXiv:1811.12808. [Online]. Available:http://arxiv.org/abs/1811.12808

[17] S. Palaniappan and R. Awang, ``Intelligent heart disease prediction systemusing data mining techniques," in Proc. IEEE/ACS Int. Conf. Comput. Syst.Appl., Mar. 2008, pp. 108115.

[18] E. O. Olaniyi, O. K. Oyedotun, and K. Adnan, ``Heart diseases diagnosisusing neural networks arbitration," Int. J. Intell. Syst. Appl., vol. 7, no. 12,p. 72, 2015.

[19] R. Das, I. Turkoglu, and A. Sengur, ``Effective diagnosis of heart diseasethrough neural

networks ensembles," Expert Syst. Appl., vol. 36, no. 4,pp. 76757680, May 2009.

[20] O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, and G. Li,``An integrated decision support system based on ANN and Fuzzy_AHPfor heart failure risk prediction," Expert Syst. Appl., vol. 68, pp. 163172,Feb. 2017.

[21] A. V. S. Kumar, ``Diagnosis of heart disease using fuzzy resolution mechanism," J. Artif. Intell., vol. 5, no. 1, pp. 4755, Jan. 2012