International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 9 Sep 2021, pp: 455-460 www.ijaem.net ISSN: 2395-5252

A Survey on Heart Disease Symptomps Prediction

Nancy Patel, Prof. Rajendra Arakh
[1]Student, Global Nature Care Sanghthan Group of Institutions, Jabalpur, MP
[2] Prof, Global Nature Care Sanghthan Group of Institutions, Jabalpur, MP

Submitted: 01-09-2021 Revised: 09-09-2021 Accepted: 12-09-2021

ABSTRACT:
Heart disease is a most harmful one that will cause death. It has a serious long term disability. This disease attacks a person so instantly. Medical data is still information rich but knowledge poor. Therefore diagnosing patients correctly on the basis of time is an exigent function for medical support. An invalid diagnosis done by the hospital leads for losing reputation. The precise diagnosis of heart disease is the dominant biomedical issue. The motivation of this paper is to study various techniques used for prediction of heart disease using machine learning techniques with neural networks that can help remedial situations.

KEYWORDS: Opinion Mining, Multi-Language, NLP, LSTM, BERT and Transformer.

I. INTRODUCTION

A heart is a vital organ of the human body. If a heart does not perform its operation properly, it will influence the other organ of the human-like kidney, brain, etc. According to the statistical data from

WHO, one-third population worldwide died from heart disease; heart disease is found to be the leading cause of death in developing countries by 2017. The heart pumps blood through the blood vessels of the circulatory system. Blood provides the body with oxygen and nutrients, as well as assisting in the removal of metabolic wastes. In the event, if blood in the body is insufficient then many organs like cerebrum suffer and if heart quits working by, death happens inside minutes. Heart disease risk factors include:

- High Cholesterol
- High blood pressure
- Diabetics
- Smoking
- Consuming too much alcohol
- Being overweight or obese
- Family history of coronary illness.

Heart disease is a grave disease that influences the heart's functionality and gives rise to complications such as infection of the coronary artery and diminished blood vessel function. Heart disease patients do not feel sick until the very last stage of the disease, and then it is too late because the damages have become irretrievable [1]. In the past years, sensor networks for healthcare IoT have advanced quickly, so it is now possible to incorporate instantaneous health data by linking bodies and sensors [2]. It is important to diagnose patients early by means of ECG. IoT-based heart attack detection systems [3] raise privacy and security concerns. As mobile devices are possible targets for malevolent attacks, more studies are needed on safety countermeasures. A fault-tolerant algorithm should be developed for a dependable IoT system [4].

Heart disease (HD) is the critical health issue and numerous people have been suffered by this disease around the world [5]. The HD occurs with common symptoms of breath shortness, physical body weakness and, feet are swollen [6]. 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 [7]. The diagnosis and treatment of heart disease is extremely difficult when modern technology and medical experts are not available [8]. The effective diagnosis and proper treatment can save the lives of many people [9].

According to the European Society of Cardiology, 26 million approximately people of HD were diagnosed and diagnosed 3.6 million annually [10]. Most of the people in the United States are suffering from heart disease [11]. 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 analyze [12]. Thus, to develop a noninvasive diagnosis system based on classifiers of machine learning (ML) to resolve these issues. Expert decision system based on machine learning classifiers and the application of artificial fuzzy logic is effectively diagnosis the HD as a result, the ratio of death decreases [13] and [14]. The Cleveland heart disease data set was used by various researchers [15] and [16] 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 improve by using proper and related features from the data. Therefore, data balancing and feature selection is significantly important for model performance improvement.

In order to improve the predictive 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 Ant Colony Optimization (ACO), fruit fly optimization (FFO), Bacterial Foraging Optimization (BFO) etc. Similarly Yun et al. [17] 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. [18] 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. Feature selection has great influence in numerous applications such as building simpler, increasing learning performance, creating clean and understandable data. The feature selection from big data is challenging job and create big problems because big data has many dimensions. Further, challenges of feature selection for structured, heterogeneous and streaming data as well as its scalability and stability issues. For big data analytics challenges of feature selection is very important to resolve. In [19] designed unsupervised hashing scheme, called topic hyper graph hashing, to report the limitations. Topic hyper graph hashing effectively mitigates the semantic shortage of hashing codes by exploiting auxiliary texts around images. The proposed Topic hyper graph hashing can achieve superior performance equalled with numerous state-of-the art approaches, and it is more appropriate for mobile image retrieval. The feature selection algorithms are classified into three types such as filter based, wrapper based and embedded based. All these feature selection mechanisms have some advantages and limitations in certain cases. The filter based method measures the relevance of a feature by correlation with the dependent variable while the wrapper feature selection algorithm measure the usefulness of a subset of features by actually training the classifier on it. The filter method is less computationally complex than wrapper method. The feature set selected by the filter is general and can be applied to any model and it is independent of a specific model.

In feature selection global relevance is of greater importance. On another hand suitable machine learning model is necessary for good results. Obviously, a good machine learning model is a model that not only performs well on data seen during training (else a machine learning model could simply learn the training data), but also on unseen data [20].

II. LITERATURE REVIEW

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 improve 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 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 Ant Colony Optimization (ACO), fruit fly optimization (FFO), Bacterial Foraging Optimization (BFO) etc.

Similarly Yun et al. 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.

Feature selection has great influence in numerous applications such as building simpler, increasing learning performance, creating clean and understandable data. The feature selection from big data is challenging job and create big problems because big data has many dimensions. Further, challenges of feature selection for structured, heterogeneous and streaming data as well as its scalability and stability issues.

For big data analytics challenges of feature selection is very important to resolve. In [15] designed unsupervised hashing scheme, called topic hyper graph hashing, to report the limitations. Topic hyper graph hashing effectively mitigates the semantic shortage of hashing codes by exploiting auxiliary texts around images. The proposed Topic hyper graph hashing can achieve superior performance equalled with numerous state-of-the-art approaches, and it is more appropriate for mobile image retrieval. The feature selection algorithms are classified into three types such as filter based, wrapper based and embedded based. All these feature selection mechanisms have some advantages and limitations in certain cases. The filter based

Method measures the relevance of a feature by correlation with the dependent variable while the wrapper feature selection algorithm measure the usefulness of a subset of features by actually training the classifier on it. The filter method is less computationally complex than wrapper method. The feature set selected by the filter is general and can be applied to any model and it is independent of a specific model. In feature selection global relevance is of greater importance.

On another hand suitable machine learning model is necessary for good results. Obviously, a good machine learning model is a model that not only performs well on data seen during training (else a machine learning model could simply learn the training data), but also on unseen data. To evaluate all classifiers on data and find that they get, on average, 50% of the cases right [16]. Furthermore, appropriate cross validation techniques and performance evaluation metrics are critical necessary for a model when model is train and test on dataset.

2.1 Machine Learning Algorithms

Classification can be described as a supervised learning algorithm in the Machine learning process. It assigns class labels to data objects based on prior knowledge of class which the data records belong. It is a Data mining technique, has made it possible to co-design and co-develop software and hardware, and hence, such components. However, integration deals with knowledge extraction from database records and prediction of class label from unknown data set of records. In classification a given set of data records is divided into training and test data sets. The training data set is used in building the classification model, while the test data record is used in validating the model. The model is used to classify and predict new set of data records that is different from both the training and test dataset. Supervised learning algorithm (like classification) is preferred to unsupervised learning algorithm (like clustering) because its prior knowledge of the class labels of data records makes feature/attribute selection easy and this leads to good prediction/classification accuracy.

Some of the common classification algorithms used in Data mining and decision support systems are: Neural networks, Logistic regression, Decision tree etc. Among these classification algorithms Decision tree algorithms is the most commonly used because it is easy to understand and cheap to implement. It provides a modeling technique that is easy for human to comprehend and simplifies the classification process. Most Decision tree algorithms can be implemented in both serial and parallel form while others can only be implemented in either serial or parallel form. Parallel implementation of decision tree algorithms is desirable inorder to ensure fast generation of results especially with the classification/prediction of large data sets; it also exploits the underlying computer architecture. But serial implementation of decision algorithm is easy to implement and desirable when small-medium data sets are involved.

S. Ambekar and R. Phalnikar [21] heart disease prediction on the basis of the dataset with help of Naïve bayes and KNN algorithm. To extend this work, we propose the disease risk prediction using structured data. We use convolutional neural network based unimodal disease risk prediction algorithm. The prediction accuracy of CNN-UDRP algorithm reaches more than 65%.

G. Suseendran et al. [22] develops a novel diagnostic system. In this paper many kinds of heart disease and symptoms responsible for that were discussed. In this paper initially heart local binary pattern and PCA have been used along with artificial neural network for heart disease prediction. Where LBP and PCA are for feature extraction and feature set size reduction and neural network for classification and verifying accuracy of the system. The complete prediction systems have been developed with Matlab tool and predict almost 95% of heart problems.

P. Ramprakash et al. [23] developed a framework in this exploration that can understand the principles of predicting the risk profile of patients with the clinical data parameters. The proposed model is constructed using Deep Neural Network and χ^2 -statistical model. The problem of under fitting and over fitting is eliminated. This model shows better results on both the testing and training data. DNN and ANN were used to analyse the efficiency of the model which accurately predicts the presence or absence of heart disease.

S. Bhoyar et al. [24] proposed a Neural Networks model using a Multilayer Perceptron (MLP) is proposed for the prediction system. Experimental analysis resulted in an accuracy of 85.71% for UCI Heart Disease dataset and 87.30% for Cardiovascular Disease dataset. When compared to previous research the increase in accuracy was approximately 12-13%. A simple web application tool is also developed using python programming to test the prediction system. This research works towards making a comprehensible tool for medical professionals as well as common people.

F. Tasnim and S. U. Habiba [25] proposed data mining classification techniques i.e. Naive Bayes (NB), Support Vector Machine (SVM), k-nearest neighbors' (k-NN), Decision Tree (DT), Neural Network (NN), Logistic Regression (LR), Random Forest (RF), Gradient Boosting are proposed to predict the probability of the coronary heart disease. In the present world, researchers are trying heart and soul to make advancements in the smart health care system. An automated system predicting the risk of heart disease may be added as a great achievement. This work of predicting heart disease is evaluated using the dataset from the UCI machine learning repository. The feature selection method enhances the performance of traditional machine learning algorithms. Among the classification algorithms, Random Forest (RF) algorithm with PCA has given the best accuracy of 92.85% for heart disease classification.

III. CONCLUSION

The various heart disease prediction techniques are discussed and analyzed in this paper. The data mining techniques used to predict heart diseases are discussed here. Heart disease is a mortal disease by its nature. This disease makes several problems such as heart attack and death. In the medical domain, the significance of data mining is perceived. Various steps are taken to apply pertinent techniques in the disease prediction. The research works with effective techniques that are done by different researchers were studied in this paper.

From the comparative study we can conclude that Neural Network technique is an efficient method for predicting heart disease. It gives good accuracy.

REFERENCES

[1].P. M. Kumar, S. Lokesh, R. Varatharajan, G. C. Babu, and P. Parthasarathy, “Cloud and IoT based disease prediction and diagnosis system for healthcare using Fuzzy neural classifier," Future Gener. Comput. Syst., vol. 86, pp. 527_534, Sep. 2018, doi: 10.1016/j.future.2018.04.036.

[2].L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan, “An automated diagnostic system for heart disease prediction based on _2 statistical model and optimally configured deep neural network," IEEE Access, vol. 7, pp. 34938_34945, 2019, doi: 10.1109/ACCESS.2019.2904800.

[3].P. K. Gupta, B. T. Maharaj, and R. Malekian, “A novel and secure IoT based cloud centric architecture to perform predictive analysis of users activities in sustainable health centres," Multimedia Tools Appl., vol. 76, no. 18, pp. 18489_18512, Sep. 2017, doi: 10.1007/s11042-016-4050-6.

[4].G. Rathee, A. Sharma, H. Saini, R. Kumar, and R. Iqbal, ``A hybrid framework for multimedia data processing in IoT-healthcare using blockchain technology," Multimedia Tools Appl., vol. 2, pp. 1_23 Jun. 2019, doi: 10.1007/s11042-019-07835-3.

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

[6].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. 255_260, 2016.

[7].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 scientific statement from the American heart association," Circulation, vol. 125, no. 15,pp. 1928_1952, 2012.

[8].S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, “Innovative artificial neural networks-based decision support system for heart diseases diagnosis," J. Intell. Learn. Syst. Appl., vol. 5, no. 3, 2013, Art. No. 35396.

[9].Q. K. Al-Shayea, “Artificial neural networks in medical diagnosis," Int. J. Comput. Sci. Issues, vol. 8, no. 2, pp. 150_154, 2011.

[10]. J. Lopez-Sendon, “The heart failure epidemic," Medicographia, vol. 33, no. 4, pp. 363_369, 2011.

[11]. 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. 933_944, 2011.

[12]. A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, “Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity," J. Roy. Soc. Interface, vol. 8, no. 59, pp. 842_855, 2011.

[13]. 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.

[14]. 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. 224_231, 2018.

[15]. R. Detrano, A. Janosi,W. Steinbrunn, M. P_sterer, 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. 304_310, Aug. 1989.

[16]. J. H. Gennari, P. Langley, and D. Fisher, ``Models of incremental concept formation," Artif. Intell. vol. 40, nos. 1_3, pp. 11_61, Sep. 1989.

[17]. Y. Li, T. Li, and H. Liu, ``Recent advances in feature selection and its applications," Knowl. Inf. Syst., vol. 53, no. 3, pp. 551_577, Dec. 2017.

[18]. J. Li and H. Liu, ``Challenges of feature selection for big data analytics," IEEE Intell. Syst., vol. 32, no. 2, pp. 9_15, Mar. 2017.

[19]. L. Zhu, J. Shen, L. Xie, and Z. Cheng, ``Unsupervised topic hyper graph hashing for efficient mobile image retrieval," IEEE Trans. Cybern., vol. 47, no. 11, pp. 3941_3954, Nov. 2017.

[20]. S. Raschka, ``Model evaluation, model selection, and algorithm selection in machine learning," 2018, arXiv: 1811.12808.

[21]. S. Ambekar and R. Phalnikar, "Disease Risk Prediction by Using Convolutional Neural Network," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1-5, doi: 10.1109/ICCUBEA.2018.8697423.

[22]. G. Suseendran, N. Zaman, M. Thyagaraj and R. K. Bathla, "Heart Disease Prediction and Analysis using PCO, LBP and Neural Networks," 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), 2019, pp. 457-460, doi: 10.1109/ICCIKE47802.2019.9004357.

[23]. P. Ramprakash, R. Sarumathi, R. Mowriya and S. Nithya vishnupriya, "Heart Disease Prediction Using Deep Neural Network," 2020 International Conference on Inventive Computation Technologies

(ICICT), 2020, pp. 666-670, doi: 10.1109/ICICT48043.2020.9112443.

[24]. S. Bhoyar, N. Wagholikar, K. Bakshi and S. Chaudhari, "Real-time Heart Disease Prediction System using Multilayer Perceptron," 2021 2nd International Conference for Emerging Technology (INCET), 2021, pp. 1-4, doi: 10.1109/INCET51464.2021.9456389.

[25]. F. Tasnim and S. U. Habiba, "A Comparative Study on Heart Disease Prediction Using Data Mining Techniques and Feature Selection," 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), 2021, pp. 338-341, doi: 10.1109/ICREST51555.2021.9331158.