
ISSN: 2321-2152
IJMECE
International Journal of modern
electronics and communication engineering
E-Mail
editor.ijmece@gmail.com
editor@ijmece.com
www.ijmece.com

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

Heart disease identification method in E-Health care using machine learning

[1]MR.CH SURESH, [2]KOPPINEEDI SURYA MOUNIKA

[1](Assistant Professor), MCA, Swarnandhra College

[23]MCA, scholar, Swarnandhra College

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

1.INTRODUCTION

An enormous number of individuals all around the globe have been affected by heart disease (HD), making it a serious health concern. Breathing difficulties, generalized weakness, and swelling feet are frequent signs of HD. Scientists are attempting to foster a more viable technique for recognizing heart disease, as existing strategies for determination are insufficient for early discovery for various reasons (counting exactness and execution time).Without today's tools and trained professionals, heart disease detection and treatment becomes a monumental challenge. Many people's lives may be spared with an accurate diagnosis and the right medications. There were almost 26 million HD diagnoses and an additional 3.6 million each year, according to the European Society of Cardiology. This condition affects the vast majority of Americans. Conventionally, a doctor would look at the patient's medical history, the results of a physical exam, and any relevant symptoms to make a diagnosis of HD. On the other hand, this diagnostic approach does not reliably detect HD patients. On top of that, it's not cheap and computationally challenging to examine.

Hence, to address these concerns, we need to create a non-invasive diagnostic system that relies on ML classiers. Artificial fuzzy logic and machine learning classifiers provide the basis of an expert judgment system that is diagnosis of HD, several researchers have used the HD data collection as a resource for studying heart disease and working to identify HD. To train and evaluate, machine learning prediction models need accurate data. On the off chance that an AI model is prepared and tried on a decent dataset,its performance will improve. Implementing relevant and appropriate data elements further enhances the model's prediction ability. Consequently, improving model performance greatly relies on data balance and feature selection. Many researchers have offered many methods for diagnosing HD in the literature, however none of these methods have shown to be useful. Data pretreatment is crucial for data standards if people want machine learning models to be more predictive. Standard scalar (SS), min-max scalar, and other preprocessing methods for removing dataset instances with missing feature values! Furthermore, the model's performance is being enhanced by the feature extraction and selection procedures. Important feature selection is typically accomplished using a variety of feature selection techniques, including LASSO, Relief, MRMR, PCA, GA, LLBFS, and optimization methods like

Anty Conley Optimization (ACO), Fruit Y Optimization (FFO), Bacterial Foraging Optimization (BFO), and others. In a similar vein, Yun et al. outlined many methods for feature selection, including secure include determination, highlight choice for enormous scope information, and component choice for high-layered little example size information. Highlights like circulated include determination, online component choice, multi-mark highlight determination, antagonistic element choice, multi-view highlight determination, and stable element choice ,were also covered. For massive data, Jindong went over the difficulties of feature selection (FS). Because of the "curse of dimensionality," some learning tasks require reducing the data's dimensionality.

Many applications benefit greatly from feature selection, including making complex structures easier to grasp, improving learning performance, and producing clean and comprehensible data. Because large data has so many dimensions, feature selection from it is a difficult task that sometimes leads to catastrophic failures. More so, there are problems with the stability and scalability of feature selection when dealing with structured, heterogeneous, and streaming data. Overcoming difficulties with feature selection is critical for big data analytics. The topic hyper graph hashing algorithm was developed as an unsupervised hashing method to report on these deficiencies. To compensate for the lack of semantic hashing codes, topic hypergraph hashing makes use of supplementary sentences around pictures. The suggested To retrieve images from mobile devices, topic hyper graph hashing is the way to go, as it outperforms several state-of-the-art methods.

Feature selection methods may be broadly categorized into three types: later, wrapper, and embedded. A variety of feature selection techniques exist, each with its own set of benefits and drawbacks. Wrapper feature selection approach assesses a subset of elements via preparing the classifier on it, however the liter-based strategy assesses highlights by relationship with the reliant variable, which is a measure of importance. Out of the two methods, the latter requires less computing power. The independent feature set chosen by the latter is generic enough to be applied to any model; it is not individual model specific. When it comes to feature selection, global relevance is really more important. However, success also requires an appropriate machine learning model. Machine learning models are considered excellent when they can effectively handle both the data viewed during training and data that has never been seen before. Using data, assess each classifier and see whether they achieve an accuracy rate of 50% or above. After a model has been trained and tested on a dataset, it is essential to use suitable cross validation procedures and performance assessment criteria. In this review, we recommended an AI based test for HD finding. With regards to distinguishing HD, AI forecast models like ANN, LR, K-NN, SVM, DT, and NB are utilized. Highlights have been picked involving cutting edge calculations as Help, m RMR, Tether, and Nearby learning-based highlights choice (LLBFS). Moreover, we recommended an elements choice methodology called quick restrictive common data (FCMIM). Finding the ideal hyperparameters for the picked model was achieved utilizing the leave-one-subject-out cross-approval (LOSO) technique. Different execution appraisal standards have been utilized for more tasteful exhibitions assessment separated from this. Tests on the Cleveland HD dataset have approved the recommended approach.

What's more, the proposed technique's presentation has been assessed in contrast with cutting edge strategies previously distributed in the writing, including NB, Triple-stage ANN (Blood vessel brain Organization) demonstrative framework, Brain network gatherings, ANN-Fluffy AHP symptomatic framework, and Versatile weighted-Fluffy framework troupe. This work adds to the current assortment of information in key ways. In the first place, the creators have a go at tackling the highlights determination issue by utilizing pre-handling methods and four cutting edge calculations: Help, MRMR, Rope, and LLBFS. They then, at that point, utilize these elements to prepare and test classiers, planning to figure out which calculations and classiers perform well with regards to exactness and preparing time. Second, to pick includes, the creators recommended the quick restrictive sharedinformation (FCMIM) FS technique. These features are then fed to the classier, which may improve prediction accuracy while decreasing computing time. Results for classifiers trained on features chosen by both the proposed FS method and state-of-the-art, conventional FS techniques are compared. Thirdly, analyze the dataset for weak attributes that impact the classier's performance. Last but not least, it claims that the FCMIM-SVM system can successfully detect HD. After this introduction, the remainder of the article is coordinated as follows. To a limited extent 2, we covered the significant writing about the point. Itemized conversation of the dataset and the numerical and hypothetical comprehension of component determination andclassification techniques is included in section 3.

2.LITERATURE SURVEY

SURVEY-1: -

Researchers have suggested a number of different AI based symptomatic methodologies for HD in the writing. To represent the meaning of the proposed work, this examination study gives a couple of current AI based demonstrative philosophies. Utilizing AI grouping draws near, Detrano et al. made a HD grouping framework with an accuracy rate of 77%. Using the features selection approach and the global evolutionary algorithm, the Cleveland dataset was analyzed. Another research by Gunadie et al. established an 80.41% accurate diagnostic system for HD categorization utilizing multi-layer Perceptron and support vector machine (SVM) algorithms. A neural network that incorporates fuzzy logic was used to construct the HD classification system by Humar et al. The accuracy rate of the categorization method was 87.4 percent. The accuracy, sensitivity, and septicity were measured at 89.01%, 80.09%, and 95.91%, respectively, by Resul et al.'s artificial neural network (ANN) ensemble-based HD diagnostic system using the statistical measurement system enterprise miner (5.2). A HD diagnostic method based on ML was created by Akil et al. The ANN-DBP algorithm, in conjunction with the FS method, yielded satisfactory results. In order to identify HD, Palaniappan et al. suggested a medical diagnostic approach that is based on expert opinion. Various machine learning prediction models, including navies bays (NB), decision trees (DT), and artificial neural networks (ANNs), were used during the system's development. The NB classifier reached an accuracy of 86.12%, the ANN classifier 88.12%, and the DT classifier 80.4%. To forecast HD in angina, Olaniyi et al. created a three-stage method based on the artificial neural network method, and it reached an accuracy of 88.89%. Samuel et al. created a medical decision support system for HD diagnosis that uses an artificial neural network and Fuzzy AHP. The accuracy performance of the suggested approach was 91.10%. A HD categorization approach based on relief and rough set techniques was suggested by Liu et al. With a classification accuracy of 92.32%, the suggested technique was successful. The authors suggested a feature-selection and classification algorithm-based HD identification technique. The SBS FS Sequential Backward Selection Algorithm is used for this purpose. We have tested the K-Nearest Neighbor (K-NN) classifier on both the whole and chosen feature sets. High accuracy was achieved by the suggested approach. Another work by MOHAN et al. [27] used mixed machine learning approaches to build a solution for HD prediction. In order to train and evaluate machine learning classifiers effectively, he also suggested a novel way to choose important features from the data. Their accuracy in categorization has been measured at 88.07%. Using an enhanced SVM-based duality optimization method, Geweid et al. [28] developed HD detection methods. To help you better understand the significance of our suggested methodology, we have summarized the limitations and benefits of the various HD diagnostic approaches in Table 1 of the aforementioned literature. In order to detect HD early on, all of these current systems required a plethora of ways. However, when it comes to predicting HD, none of these methods are very accurate and need a lot of computing power. Table 1 shows that in order to increase therapy and recovery, the HD detection method's prediction accuracy requires more work for every client. Low accuracy and excessive calculation time were the main problems with these earlier methods, which may have been caused by using irrelevant characteristics in the dataset.

Survey 2:

The second survey included a method for detecting cardiac disease using machine learning and sampling strategies to deal with imbalanced datasets. Random Over-Sampling, Synthetic Minority Over-Sampling (SMOTE), and Adaptive Synthetic Sampling Approach (ADASYN) are some of the sampling strategies that are used. In order to train and test the algorithms, we used the Framingham datasets available on the Kaggle website. These datasets comprise 4239 instances with a total of 15 characteristics. The objective was to determine a patient's 10-year risk of developing coronary heart disease using the attributes. Learning representations (LR, KNN, AdaBoost, DT, NB, and RF) are used in the machine learning process. Accuracy, recall, and precision were the metrics used to evaluate these categorization systems. Different sample methods call for different values for each of these factors. Their testing findings showed that the SVM classifier using the Random Over-Sampling approach had the highest accuracy (99%) in predicting the occurrence of heart disease. Nevertheless, RF outperformed DT classifier and ADASYN with an accuracy of 90.3% when using the SMOTE approach.

3. EXISTING SYSTEM

Utilizing AI characterization draws near, Detrano made a HD order framework with an exactness pace of 77%.Using the features selection approach and the global evolutionary algorithm, the Cleveland dataset was analyzed. Guddle obtained an accuracy of 80.41% in a different research he conducted using a diagnostic system he built utilizing support vector machine (SVM) and multi-layer Perceptron algorithms for HD categorization. Using a neural network that included fuzzy logic, Humar developed an HD categorization system. The accuracy rate of the categorization method was 87.4 percent. Using enterprise miner, a statistical measurement system, and an artificial neural network (ANN) ensemble, Resul created an HD diagnostic system with an accuracy of 89.01%, sensitivity of 80.09%, and specificity of 95.91%. A HD diagnostic method based on ML was created by Akil et al. [24]. The ANN-DBP algorithm, in conjunction with the FS method, yielded satisfactory results.

In order to identify HD, Palaniappan suggested a medical diagnosing system that is run by experts. Various machine learning prediction models, including navies bays (NB), decision trees (DT), and artificial neural networks (ANNs), were used during the system's development. The NB classifier reached an accuracy of 86.12%, the ANN classifier 88.12%, and the DT classifier 80.4%.

In order to forecast HD in angina, Olaniyi created a three-stage method based on the artificial neural network method, and it arrived at an exactness of 88.89%. Samuel fabricated a clinical choice emotionally supportive network for HD finding utilizing a fake brain organization and Fluffy AHP. The precision execution of the recommended approach was 91.10%. Liu recommended a technique for HD order that utilizes help and unpleasant set strategies. With a grouping precision of 92.32%,the suggested technique was successful. The authors suggested a feature-selection and classification algorithm-based HD identification technique. The SBS FS Sequential Backward Selection Algorithm is used for this purpose. Both the whole and partial feature sets were used to evaluate the K-Nearest Neighbor (K-NN) classifier. High accuracy was achieved by the suggested approach. Another work by MOHAN used mixed machine learning approaches to create a solution for HD prediction. In order to train and evaluate machine learning classifiers effectively, he also suggested a novel way to choose important features from the data. Their accuracy in categorization has been measured at 88.07%.

Geweid developed HD identifying methods by enhancing a duality optimization strategy based on support vector machines. Table 1 summarizes the benefits and drawbacks of the suggested HD diagnostic approaches from the aforementioned literature, which should help readers to grasp the significance of our approach. In order to detect HD early on, all of these current systems required a plethora of ways. While these methods can forecast HD, they aren't very accurate, and they need a lot of computing power.

3.1 PROPOSED SYSTEM

In this study, we provide a device that can detect HD using an approach based on machine learning. For HD identification, machine learning prediction models are used, including ANN, LR, K-NN, SVM, DT, and NB. When choosing these features, we consulted the gold standard in feature selection algorithms like Relief, MRMR, LASSO, and Local-learning-based features-selection (LLBFS).

Additionally, the system suggested an approach for features selection called fast conditional mutual information (FCMIM). For optimal model selection, the leave-one-subject-out cross-validation (LOSO) method was used to choose the hyper-parameters. The evaluation of classifier performance has also made use of many performance assessment indicators. An evaluation of the suggested approach was conducted using the

Cleveland HD dataset. In addition, the suggested method's performance has been evaluated in comparison to state-of-the-art methods already published in the literature, including NB, a three-phase ANN (Artificial neural Network) diagnosis system [18], Neural network ensembles (NNE), an ANN-Fuzzy-AHP diagnosis system (AFP), and an Adaptive-weighted-Fuzzy-system-ensemble (AWFSE). Contributions to the field from this work are as follows.

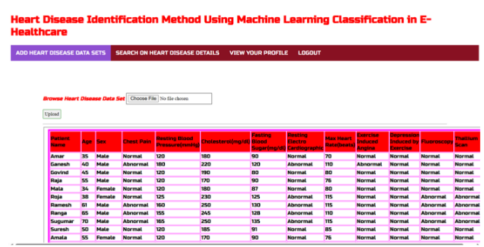
The authors begin by attempting to solve the highlights determination issue by utilizing pre-handling strategies and four cutting edge calculations: Alleviation, MRMR, Rope, and LLBFS.They then use this subset of features to train and test classifiers effectively, determining which algorithm and classifier performs well in terms of accuracy and training time.

After that, in order to choose features, the authors suggested the fast conditional mutual information (FCMIM) FS technique. Classifiers then use these features to make better predictions with less computing time. The suggested FS algorithm's chosen characteristics were compared to those of the conventional state-of-the-art FS algorithms in order to assess the classifiers' performances.

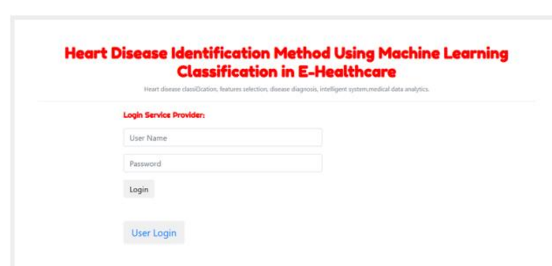
Lastly, find the dataset's weak characteristics that impact the classifiers' performance. Lastly, it is suggested that the FCMIM-SVM system might successfully detect HD.

4. OUTPUT SCREENS

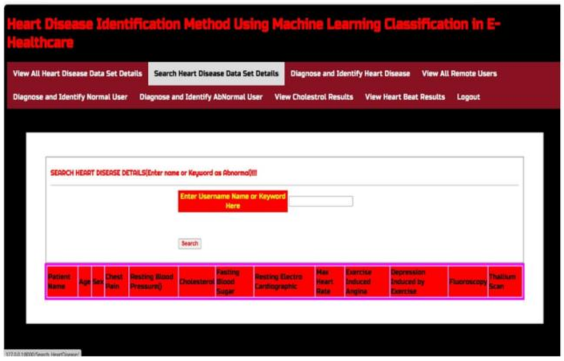
User login:



Service Provider Login Page



ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024



5. CONCLUSION

The purpose of this research was to create a system for the accurate identification of cardiac illness using machine learning. Machine learning classifiers such as LR, K-NN, ANN, SVM, NB, and DT were used in the system's construction. In order to address the feature selection issue, we used four well-established algorithms: Relief, MRMR, LASSO, and LLBFS. Additionally, we introduced a new approach, FCMIM. The system use the LOSO cross-validation approach to select the optimal hyperparameters. The Cleveland heart disease dataset is used for system testing. In addition, the identification system's performance is checked using performance assessment measures. Table 15 shows that out of the four feature selection techniques tested (MRMR, LASSO, LLBFS, and FCMIM), the ANN classifier performs the best on the Relief FS algorithm in terms of septicity. Consequently, the most effective method for predicting the presence of healthy individuals is ANN with relief. Classier NB's sensitivity to features chosen by the LASSO FS method outperforms Relief FS's sensitivity to features chosen by the classier SVM (linear) algorithm. Applying the FCMIM FS technique to a subset of features yields a 91% classier Logistic Regression MCC. When compared to MRMR FS algorithms and others, Logistic Regression with Relief, LASSO, FCMIM, and LLBFS FS have the best processing time. Thus, in comparison to the conventional feature selection methods, the experimental findings demonstrate that the suggested approach achieves high classification accuracy by selecting characteristics that are more effective. The most relevant and crucial traits, as determined by feature selection algorithms, are exercise-induced angina and chest discomfort of the Thallium Scan kind. The trait is shown by all FS algorithm outcomes. The diagnosis of cardiac disease should not be based on fasting blood sugar (FBS). Table 17 shows that the suggested feature selection technique (FCMIM) achieves an SVM accuracy of 92.37%, which is excellent when compared to other offered approaches. Additionally, when it comes to HD detection, the performance of the machine learning-based method FCMIMSVM is higher than that of the deep neural network. In the diagnosis of life-threatening illnesses, even a little gain in prediction accuracy may have a huge impact. The innovative aspect of the research is creating a method for detecting cardiac problems. In this research, characteristics are selected using five different algorithms: four traditional ones and one suggested one. We employ the LOSO CV technique and performance indicators to measure success. For the purpose of testing, the Cleveland heart disease dataset is used. Our research leads us to believe that a decision-support system powered by machine learning algorithms would be better suited to the task of diagnosing cardiac illness. We also know that superfluous characteristics increase calculation time and hurt the diagnosing system's performance. Hence, our study's use of features selection algorithms to choose the right characteristics to boost classification accuracy and speed up the diagnostic system's processing time is another novel aspect. To improve the prediction system's effectiveness for HD diagnosis, we will include more characteristics selection algorithms and optimization methodologies in the future. I will continue to focus on the treatment and recovery of illnesses in the future, including essential ones like heart, breast, Parkinson's, and diabetes, since after diagnosis, managing and treating the condition becomes extremely important.

6.REFERENCES

• L. Bui, T. B. Horwich, and G. C. Funaro, ``Epidemiology and risk prole of heart failure," Nature Rev. Cardio, vol. 8, no. 1, p. 30, 2011.

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.

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,

S. Ghanem, A. Mohammad, and A. Al-Ibrahim, ``Innovative artificial

neural networks-based decision support
system for heart diseases diagnosis,"
J. Intel. Learn. Syst. Appl., vol. 5, no. 3,
2013, Art. no. 35396.

Q. K. Al-Shaye, ``Arterial neural networks
in medical diagnosis," Int. J.
Compute. Sci. Issues, vol. 8, no. 2, pp.
150154, 2011. J. Lopez-Sendon, ``The heart
failure epidemic," Musicographic, vol. 33,
P. A. Heidenreich, J. G. Trogdon, O. A.
Khavjou, J. Butler, K. Dracup, M. D.
Meyerowitz, E. A. Finkelstein, Y. Hong, S.
C. Johnston, A. Khera,

D. M. Lloyd-Jones, S. A. Nelson, G. Nichol,
D. Orenstein, P.W. Wilson,

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.
Tsana, M. A. Little, P. E. McSharry, and L.
O. Ramig, ``Nonlinear speech analysis
algorithms mapped to a standard metric
achieve clinically.