Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 2, No. 2, February : 2024

AN OPTIMAL ML APPROACH TO FEATURE ENGINEERING FOR THE PREDICTION OF HEART DISEASES

G Madhu, Assistant . Professor, Dept. of CSE Dept. , MVSR Engg College
Dr. Akhil Khare, Professor, Dept. of CSE Dept. , MVSR Engg College

ABSTRACT:
Heart failure is a chronic condition that impacts a significant global population. There is a pressing need for the development of an effective machine learning-based approach to accurately forecast the health condition of individuals afflicted with heart failure at an early stage. This would enable timely intervention and mitigation strategies to address this pervasive global issue. Pharmaceutical intervention remains the primary therapeutic approach for heart failure management; however, there is an increasing acknowledgment within the medical community of the efficacy of exercise as an adjunctive modality in the treatment of this condition. This study employed machine learning techniques to devise a methodology for enhancing the diagnosis of heart failure by leveraging patient health indicator data. Our research endeavors facilitate the identification of heart failure during its initial phases, thereby contributing to the preservation of human lives. In this study, a total of nine machine learning methods were employed for comparative analysis. Additionally, a novel feature engineering technique, known as Principal Component Heart Failure (PCHF), was proposed to identify and prioritize the most significant features with the aim of enhancing performance. A novel set of features was proposed as a means to enhance the suggested PCHF system and optimize accuracy scores. The new dataset was constructed by utilizing the eight most suitable traits. A series of experiments were conducted to assess the efficacy of various methodologies. The decision tree method recommended in this study outperformed the machine learning models employed and other contemporary research endeavors, achieving a remarkable accuracy score of 100%, which is noteworthy. The efficacy of all employed methods was demonstrated through the utilization of the cross-validation technique. The proposed research study is expected to contribute significant scientific advancements to the field of medicine.
Keywords *– Machine learning, heart failure, cross validations, feature engineering*

1. INTRODUCTION

Heart failure is a situation in which the heart can't pump enough blood to meet the body's needs [1]. Cardiovascular illnesses have become a major global health issue that has a big effect on people's health all over the world. Heart failure is a common and dangerous disease that affects millions of people all over the world. A new report says that heart failure diseases affect about 26 million people [2]. There are two types of things that can cause heart failure. First, it had to do with the shape of the heart, like a past heart attack. Second, it has to do with how the heart works, like excess blood pressure. Heart failure can cause shortness of breath, tiredness, and swollen feet and legs. Heart failure can be treated with medicines, changes in living, and, in some cases, surgery. Research has shown that finding and treating heart failure early can improve the quality of life and make people live longer [3]. The main goal of the current study is to create a machine-learning model for controlling heart failure in order to improve the health of patients. Medical findings and the health care business use machine learning a lot [4]. Machine learning has many uses in the medical field, such as finding new drugs, figuring out what's wrong with an image, predicting outbreaks, and figuring out when someone will have a heart attack. Large amounts of medical data can be used to help machines learn trends and do prediction analysis. When compared to traditional medical methods, machine learning has many benefits. For example, it saves time and money, which helps improve analysis.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 2, No. 2, February : 2024

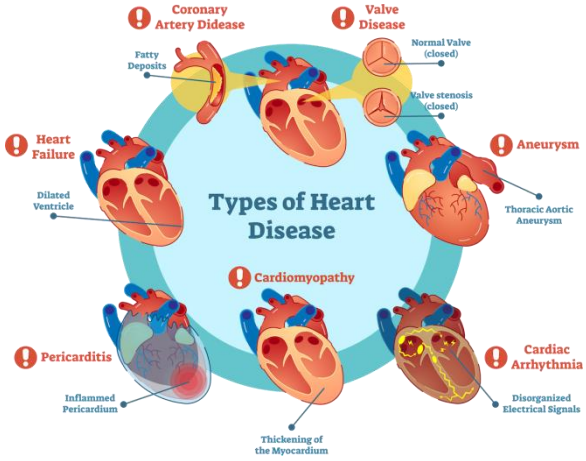

Fig.1: Example figure

To improve speed, a new PCHF feature engineering method is suggested to choose the most important features. Using the suggested PCHF technique, eight high-importance features from the dataset are chosen to build machine learning methods. We made an improvement to the suggested PCHF mechanism by adding a new set of features. This allowed us to get the best accuracy scores of any proposed method. • The comparison uses the nine advanced models of machine learning to predict heart failure. For each machine learning method used, the hyperparameters tuning is done to find the best-fit parameters and get a high-performance accuracy score. We used the k-fold cross-validation method to check the success of machine learning models that were already in use.

2. LITERATURE REVIEW

Chronic Heart Failure Detection from Heart Sounds Using a Stack of Machine-Learning Classifiers:

Chronic heart failure is a broad sickness that influences in excess of 26 million individuals all over the planet. It is a central motivation behind why such countless individuals with coronary illness pass on, and it sends in excess of 1,000,000 individuals to the emergency clinic consistently in Europe and North America. Strategies for recognizing persistent cardiovascular breakdown can be utilized to act preventively, work on early ID, and stay away from hospitalisations or even dangerous circumstances, which enormously works on the personal satisfaction for the patient. In this review, we portray a method for utilizing ML to sort out whether or not somebody has ongoing cardiovascular breakdown by paying attention to their heart sounds. Sifting, sectioning, highlight extraction, and ML are portions of the technique. A leave-one-subject-out scoring technique was involved on information from 122 subjects in the review to test the cycle. The strategy was right 96% of the time, which was 15 rate focuses better compared to a larger part indicator. All the more explicitly, it views as 87% individuals with persistent cardiovascular breakdown with an exactness of 87%. The review demonstrated that cutting-edge ML can be utilized to find constant cardiovascular breakdown by utilizing genuine sounds got with a computerized stethoscope that doesn't disrupt everything.

Global Public Health Burden of Heart Failure:

Heart failure (HF) is an inescapable plague that effects no less than 26 million individuals and is turning out to be more normal. Medical care costs for HF are high and will rise enormously as the populace ages. Despite the fact that medicines and security have made considerable progress, demise and ailment are as yet normal and personal satisfaction is low. The expressed recurrence, occurrence, mortality, and disease rates change by area in light of the various reasons for HF and the clinical highlights of the people who have it. In this review, we center around the worldwide the study of disease transmission of HF, giving data about its recurrence, rate, passings, and ailments all over the planet.

Secure and Robust Machine Learning for Healthcare: A Survey:

Machine learning (ML) and deep learning (DL) strategies have become exceptionally well known lately on the grounds that they work better compared to different techniques for an extensive variety of medical services applications, from anticipating heart failure from one-layered heart signs to computer-aided diagnosis (CADx) utilizing multi-faceted clinical pictures. Despite the fact that ML/DL performs well, there are still inquiries regarding how well it will function in medical care settings, which are generally hard in light of the numerous security and protection issues included. This is particularly evident since late outcomes have shown that ML/DL are helpless against assaults from an external perspective. In this review, we give a layout of the various areas of medical care that utilization these techniques and show the security and protection gives that accompany them. We likewise discuss potential ways of ensuring that ML utilized in medical services applications is protected and doesn't disregard individuals' security. In conclusion, we discuss the issues confronting concentrate on this moment and a few confident regions for future exploration.

Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers:

Coronary heart diseases is one of the principal reasons individuals bite the dust everywhere. Quite possibly of the hardest thing to do in the space of clinical information examination is to anticipate coronary illness. Machine learning (ML) can assist with diagnostics by deciding and making expectations in light of information from the medical services industry all over the planet. We have likewise seen ML strategies used to recognize illnesses in the clinical region. Along these lines, many exploration papers have demonstrated the way that a ML calculation can be utilized to recognize coronary illness. In this review, we utilized eleven ML models to find the main attributes, which made coronary illness more straightforward to foresee. A few component coordinates and notable grouping strategies were utilized to flaunt the figure model. With slope helped trees and multi-facet perceptron, we had the option to distinguish coronary illness with a 95% achievement rate. With a 96% achievement rate, the Random Forest improves at of foreseeing coronary illness.

Machine Learning Techniques for Heart Disease Prediction: A Comparative Study and Analysis:

Individuals are so occupied with work and different things in their everyday lives that they disregard their wellbeing. Consistently, an ever increasing number of individuals become ill since they are excessively occupied and couldn't care less about their wellbeing. Additionally, most of individuals have an infection like coronary illness. As per figures from the World Health Organisation (WHO), coronary illness kills practically 31% of individuals all over the planet. Thus, knowing regardless of whether coronary illness will happen is significant for the clinical region. Yet, how much information that emergency clinics and the clinical field get is enormous to the point that sorting out what everything means is in some cases hard. Utilizing ML strategies to create these expectations and manage information can assist clinical experts with taking care of their responsibilities much better. Thus, in this review, we've discussed coronary illness and what causes it, as well as how ML works. We had the option to anticipate coronary illness utilizing these ML techniques. We likewise did an examination of the calculations for ML that were utilized in the conjecture work out. The general purpose of this study is to sort out some way to utilize ML to recognize coronary illness and afterward check the outcomes out.

3. METHODOLOGY

Heart disease is believed to be the most risky and lethal sickness in people, as per concentrates on that have been finished previously. Heart infections are killing an ever increasing number of individuals, which is a major issue for medical care administrations all over the planet. The vast majority of individuals who get this perilous infection are kids. In this review, the Cleveland information was utilized to further develop the way that coronary illness can be anticipated by

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 2, No. 2, February : 2024

utilizing a strategy called "highlight determination." This method assists with getting a precision of 86.60%. Finally, past examinations have found enormous openings in the exploration, which recommends that the accuracy of the exhibition isn't adequate. Thus, in this part, we cautiously check out at the presentation information from the last review. This part on connected work depends on results that sum up the viability of all models that have been utilized previously. Various kinds of models actually give different forecast scores, as per concentrates on that have been finished previously. In this way, diminishing the quantity of aspects and specialized elements can further develop the information determination, prompting more precise expectations.

Drawbacks:

1. Past examinations have found large openings in the exploration, which recommends that the rightness of the presentation doesn't depend on par.
2. Different sorts of models actually give different expectation scores, as per concentrates on that have been finished previously.

In this review, we utilized ML and wellbeing pointer information from patients to concoct a method for further developing how cardiovascular breakdown is found. Our review assists individuals with living longer by making it simpler to find cardiovascular breakdown from the beginning. We utilized another Principal Component Heart Failure (PCHF) highlight designing technique to pick the main elements and further develop execution so we could investigate nine ML based calculations. To get the best grades for exactness, we worked on the proposed PCHF process by concocting another arrangement of elements. The new assortment is based on the eight best-fit characteristics.

Benefits:

1. We did a great deal of tests to figure out how well various strategies functioned.
2. The proposed decision tree strategy showed improvement over the applied ML models and other best in class review, getting a high exactness score of 100 percent, which is noteworthy.

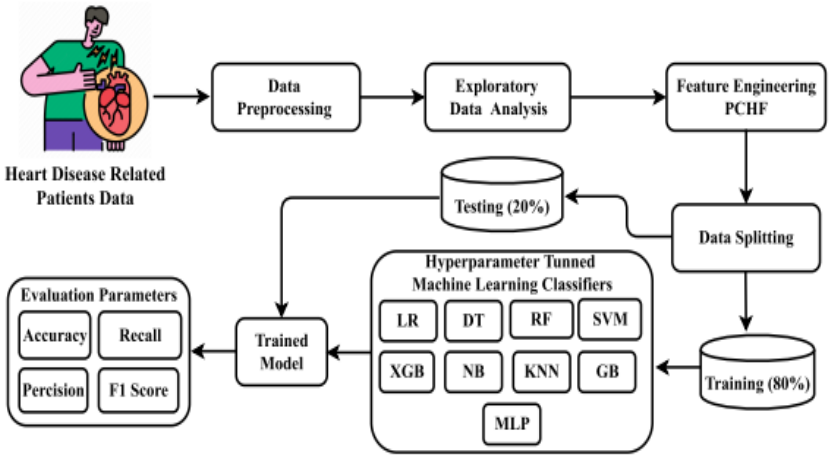

Fig.2: System architecture

MODULES:

To do the undertaking referenced above, we've made the accompanying modules:

- With this module, we will place information into the framework for information revelation.
- With this module, we will peruse information for handling.
- Dividing information into train and test: This instrument will be utilized to divide information into train and test.
- Model age: LR, DT, RF, SVM, KNN, MLP, NB, XGBoost, Gradient Boosting, Stacking Classifier(RF+DT+LightGBM) are utilized to assemble the model. Determined accuracy of calculations.
- Joining and signing in as a client: Utilizing this device will get you enlistment and signing in.
- Utilizing this device will give forecasts more data.

- Estimate: the last gauge was shown

4. IMPLEMENTATION

We secondhand the following arrangements in this place project.

LR: This somewhat mathematical model, that is otherwise known as a "logit model," is frequently secondhand for classifying data and making forecasts. Logistic regression is a habit to resolve how likely it is that entity will occur, like vote a suggestion of correction vote, established a set of free determinants.

DT: A decision tree is a non-parametric directed learning plan that maybe secondhand for both categorization and reversion questions. It is start like a timber accompanying a root bud, arms, knots in the arms, and leaf knots at greatest amount of the arms.

RF: Leo Breiman and Adele Cutler fictitious the Random Forest machine learning pattern. It uses the results of differing resolution shrubs to receive a alone answer. It has enchant cause it is handy and maybe used to answer both categorization and reversion questions.

SVM: The Support Vector Machine is a strong led treasure that everything best on limited datasets that are difficult. Support Vector Machine, or SVM, maybe secondhand for both reversion and categorization, but private cases, it everything best for categorization.

KNN: The k-nearest neighbours design, otherwise known as KNN or k-NN, is a non-parametric, directed learning prophet that uses nearness to categorize or guess how a sole dossier point fits into a group.

MLP: A multilayer perceptron (MLP) is a term for a up-to-date feedforward affected interconnected system, that is containing sufficiently related neurons accompanying a nonlinear incitement function, organized in not completely three coatings, and is popular for being capable to distinguish dossier that can't be divided in a uninterrupted habit. The name is wrong cause the original perceptron secondhand a Heaviside step function a suggestion of correction the nonlinear incitement function that current networks use.

NB: Naive Bayes Classifier is individual of the most natural and most direct categorization designs. It helps build fast machine intelligence models that can form keen forecasts. It is a probabilistic classifier, that resources that it create forecastings established how likely entity is.

XGBoost: XGBoost is a distributed gradient-boosting library namely optimised for preparation machine intelligence models fast and considerably. It is a type of knowledge named "ensemble education," that takes the results of various feeble models and puts bureaucracy together to form a better estimate.

GB: Gradient Boosting is a popular boosting design secondhand for categorization and reversion questions in machine learning. Boosting is a type of ensemble knowledge at which point the model is prepared in a row, and each new model tries to fix the mistakes fashioned for one model before it. It turns a group of feeble learners into a group of excellent learners.

Stacking Classifier: A stacking classifier is an ensemble learning pattern that blends diversified categorization models into individual "excellent" model. This can frequently bring about better act, because the combined model can get or give an advantage each model's benefits.

5. CONCLUSION

In this study, it is suggested that machine learning methods could be used to predict heart failure. The collection, which is made up of information from 1025 patients, is used to build the models that are used. We suggest a new PCHF feature engineering method that picks the eight most important features to improve speed. Among the machine learning methods that have been used, you can compare logistic regression, random forest, support vector machine, decision tree, extreme gradient boosting, naive base, k-nearest neighbours, multilayer perceptron, and gradient boosting. The suggested DT method was 100% accurate and only took 0.005 seconds to run. Each learning model's

success is checked using the cross-validation method, which is based on 10 times of data. Our proposed method did better than the best studies to date and can be used to find heart failure anywhere.

REFERENCES

1) M. Gjoreski, M. Simjanoska, A. Gradišek, A. Peterlin, M. Gams, and G. Poglajen, ‘‘Chronic heart failure detection from heart sounds using a stack of machine-learning classifiers,’’ in Proc. Int. Conf. Intell. Environments (IE), Aug. 2017, pp. 14–19.
2) G. Savarese and L. H. Lund, ‘‘Global public health burden of heart failure,’’ Cardiac Failure Rev., vol. 3, no. 1, p. 7, 2017.
3) E. J. Benjamin et al., ‘‘Heart disease and stroke statistics—2019 update: A report from the American heart association,’’ Circulation, vol. 139, no. 10, pp. e56–e528, 2019.
4) A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, ‘‘Secure and robust machine learning for healthcare: A survey,’’ IEEE Rev. Biomed. Eng., vol. 14, pp. 156–180, 2021.
5) C. A. U. Hassan, J. Iqbal, R. Irfan, S. Hussain, A. D. Algarni, S. S. H. Bukhari, N. Alturki, and S. S. Ullah, ‘‘Effectively predicting the presence of coronary heart disease using machine learning classifiers,’’Sensors, vol. 22, no. 19, p. 7227, Sep. 2022.
6) R. Katarya and S. K. Meena, ‘‘Machine learning techniques for heart disease prediction: A comparative study and analysis,’’ Health Technol., vol. 11, no. 1, pp. 87–97, Jan. 2021.
7) P. Rani, R. Kumar, N. M. O. S. Ahmed, and A. Jain, ‘‘A decision support system for heart disease prediction based upon machine learning,’’ J. Reliable Intell. Environments, vol. 7, no. 3, pp. 263–275, Sep. 2021.
8) N. S. Mansur Huang, Z. Ibrahim, and N. Mat Diah, ‘‘Machine learning techniques for early heart failure prediction,’’ Malaysian J. Comput. (MJoC), vol. 6, no. 2, pp. 872–884, 2021.
9) T. Amarbayasgalan, V. Pham, N. Theera-Umpon, Y. Piao, and K. H. Ryu, ‘‘An efficient prediction method for coronary heart disease risk based on two deep neural networks trained on well-ordered training datasets,’’ IEEE Access, vol. 9, pp. 135210–135223, 2021.
10) R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, ‘‘Prediction of heart disease using a combination of machine learning and deep learning,’’ Comput. Intell. Neurosci., vol. 2021, pp. 1–11, Jul. 2021.
11) S. Alotaibi, ‘‘Implementation of machine learning model to predict heart failure disease,’’ Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 6, pp. 1–8, 2019.
12) D. K. Plati, E. E. Tripoliti, A. Bechlioulis, A. Rammos, I. Dimou, L. Lakkas, C. Watson, K. McDonald, M. Ledwidge, R. Pharithi, J. Gallagher, L. K. Michalis, Y. Goletsis, K. K. Naka, and D. I. Fotiadis, ‘‘A machine learning approach for chronic heart failure diagnosis,’’ Diagnostics, vol. 11, no. 10, p. 1863, Oct. 2021.
13) A. Saboor, M. Usman, S. Ali, A. Samad, M. F. Abrar, and N. Ullah, ‘‘A method for improving prediction of human heart disease using machine learning algorithms,’’ Mobile Inf. Syst., vol. 2022, pp. 1–9, Mar. 2022.
14) S. Sarah, M. K. Gourisaria, S. Khare, and H. Das, ‘‘Heart disease prediction using core machine learning techniques—A comparative study,’’ in Advances in Data and Information Sciences. Springer, 2022, pp. 247–260.
15) C. Trevisan, G. Sergi, and S. Maggi, ‘‘Gender differences in brain-heart connection,’’ Brain and Heart Dynamics. 2020, pp. 937–951.
16) M. S. Oh and M. H. Jeong, ‘‘Sex differences in cardiovascular disease risk factors among Korean adults,’’ Korean J. Med., vol. 95, no. 4, pp. 266–275, Aug. 2020.
17) D. C. Yadav and S. Pal, ‘‘Prediction of heart disease using feature selection and random forest ensemble method,’’ Int. J. Pharmaceutical Res., vol. 12, no. 4, pp. 56–66, 2020.

18) D. Tomar and S. Agarwal, ‘‘A survey on data mining approaches for healthcare,’’ Int. J. Bio-Sci. Bio-Technol., vol. 5, no. 5, pp. 241–266, Oct. 2013.
19) S. Ekiz and P. Erdogmus, ‘‘Comparative study of heart disease classification,’’ in Proc. Electr. Electron., Comput. Sci., Biomed. Eng. Meeting (EBBT), Apr. 2017, pp. 1–4.
20) B. A. Tama, S. Im, and S. Lee, ‘‘Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble,’’ BioMed Res. Int., vol. 2020, pp. 1–10, Apr. 2020.
21) V. Ramalingam, A. Dandapath, and M. K. Raja, ‘‘Heart disease prediction using machine learning techniques: A survey,’’ Int. J. Eng. Technol., vol. 7, no. 2, pp. 684–687, 2018.
22) Heart Disease Dataset|Kaggle, DAVID LAPP, Atlanta, Georgia, 1988.
23) K. Shah, H. Patel, D. Sanghvi, and M. Shah, ‘‘A comparative analysis of logistic regression, random forest and KNN models for the text classification,’’ Augmented Human Res., vol. 5, no. 1, pp. 1–16, Dec. 2020.
24) M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. W. Quinn, and M. A. Moni, ‘‘Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison,’’ Comput. Biol. Med., vol. 136, Sep. 2021, Art. no. 104672.
25) A. Raza, K. Munir, M. Almutairi, F. Younas, and M. M. S. Fareed, ‘‘Predicting employee attrition using machine learning approaches,’’ Appl. Sci., vol. 12, no. 13, p. 6424, Jun. 2022.
26) A. Raza, H. U. R. Siddiqui, K. Munir, M. Almutairi, F. Rustam, and I. Ashraf, ‘‘Ensemble learning-based feature engineering to analyze maternal health during pregnancy and health risk prediction,’’ PLoS ONE, vol. 17, no. 11, Nov. 2022, Art. no. e0276525.
27) S. Shabani, S. Samadianfard, M. T. Sattari, A. Mosavi, S. Shamshirband, T. Kmet, and A. R. Várkonyi-Kóczy, ‘‘Modeling pan evaporation using Gaussian process regression K-nearest neighbors random forest and support vector machines; comparative analysis,’’ Atmosphere, vol. 11, no. 1, p. 66, Jan. 2020.
28) N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, ‘‘HDPM: An effective heart disease prediction model for a clinical decision support system,’’ IEEE Access, vol. 8, pp. 133034–133050, 2020.
29) S. Bashir, Z. S. Khan, F. H. Khan, A. Anjum, and K. Bashir, ‘‘Improving heart disease prediction using feature selection approaches,’’ in Proc. 16th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST), Jan. 2019, pp. 619–623.
30) D. D. Rufo, T. G. Debelee, A. Ibenthal, and W. G. Negera, ‘‘Diagnosis of diabetes mellitus using gradient boosting machine (LightGBM),’’ Diagnostics, vol. 11, no. 9, p. 1714, Sep. 2021.