**A Comprehensive Literature Review on ML (Machine Learning) Based Cardiovascular Disease Classification**

**Introduction:**

These days, cardiovascular disease (CVD) is regarded as a prevalent and severe medical condition. The World Health Organization (WHO) reports that noncommunicable diseases, including CVD, were the cause nearly forty million fatalities in the last 10 years. Around a third of these fatalities happened in groups with low or middle incomes. The Centers for Disease Control and Prevention (CDC) revealed statistics indicating that cardiovascular disease accounts for one out of every four deaths in the United States, or around 610,000 deaths annually. Multiple factors, such as high blood pressure, excessive alcohol intake, cholesterol, stress, and obesity, are crucial [6].

Abnormalities of the heart and blood arteries collectively are referred to as cardiovascular diseases, or CVDs. These consist of:   
  
illness of the blood arteries supplying the heart muscle is known as coronary heart disease;   
  
illness of the blood arteries supplying the brain, known as cerebrovascular disease;   
  
illness of the blood arteries supplying the arms and legs is known as peripheral arterial disease;   
  
rheumatic heart disease: streptococcal bacteria-induced rheumatic fever that damages the heart muscle and heart valves;   
  
congenital heart disease, which are birth defects resulting from structural cardiac anomalies that impair the heart's normal development and function; and   
  
Pulmonary embolism and deep vein thrombosis are blood clots in the legs that have the potential to travel to the heart and lungs [13].

Institutions gather a large quantity of data about disease diagnostic testing, patients, and many other concerns nowadays. Machine learning is a set of techniques for detecting hidden patterns or correlations in data. As a result, and machine learning algorithm is being suggested within that study for such invention of a prediction of heart disease, which has been confirmed and uses fixed automat diagnostic samples. Machine learning approaches help healthcare professionals diagnose diseases more accurately and quickly in the initial stages [9]. Over the last five years, there has been significant development in machine learning and artificial intelligence (AI). Researchers in the medical field are becoming more involved in AI systems. For instance, according to a recent survey, nearly fifty percent of the healthcare institutions surveyed using artificial intelligence in imaging or aim to do so [10].

The convergence of machine learning, artificial intelligence, and healthcare heralds a promising era in disease diagnosis, prediction, and treatment. With continued advancements in technology and data-driven approaches, there is optimism for improved patient outcomes and more efficient healthcare delivery systems. However, ethical considerations, data privacy concerns, and the need for rigorous validation remain critical areas of focus as we navigate the complexities of implementing these innovative technologies in clinical practice.

Literature Review:

Reference [1] used the UCI library's Cleveland dataset to illustrate the use of a hybrid machine learning technique for cardiovascular disease prediction.   
  
Using the combination of Random Forest and Linear Methods, the HRFLM classification strategy achieved a high accuracy rate in cardiovascular disease prediction. When compared to other techniques, the HRFLM classification algorithm performed better in terms of predicting cardiovascular disease.   
  
The evaluation metrics included overall error rate, overall classification error rate, accuracy, F-measure, sensitivity, and specificity. The HRFLM method's accuracy of 86.8% for F-measure suggests that it was effective in predicting cardiovascular disease in contrast to other strategies.

Reference [2] uses ML classification algorithms on a dataset whose target variable is categorical with categories "normal" and "diseased". The paper uses classifiers to estimate heart disease risk. After data cleaning, in which mean was used instead of missing values, EDA was used to assess the relationship between different attributes. After applying a number of classifiers, SVM achieved the highest accuracy of 96%, yielding the best results. RF showed 97% accuracy, DT and LR both fared well, with accuracies of 90% and 92%, respectively. Furthermore, 94% of the time, Techniques such as LR, MARS, EVF, and CART-ML accurately identified the presence of cardiovascular disease together. With accuracy rates ranging between approximately 95% and 97%, the study shown the overall efficacy of machine learning models in early identification and avoidance of coronary heart disease. In particular, Random Forest, Gradient Boosting Tree, and Multilayer Perceptron proved to be highly effective.

Reference [3] Using the SVM, LR, and NB algorithms, a study designed a web-based tool with the objective of predicting heart conditions using machine learning. The UCI repository's data sets, which were separated into training and testing sets, were used. Patients can enter their medical information into the application's user-friendly interface, which stores the information in a MySQL database. SVM proved to be the most accurate classifier following extensive preprocessing, reaching up to 64.4% accuracy, while LR and NB achieved 61.45% and 60% accuracy, respectively. The study emphasizes the value of innovation for improved patient outcomes while developing machine learning in healthcare, despite limits in early detection.

Reference [4]

In order to forecast coronary artery disorders (CAD), this study examines patient data analysis. It makes use of a 282 instance dataset with 58 features that was gathered from a clinic and processed to guarantee data quality by eliminating outliers and missing values. When clustering patients, descriptive methods like the K-means algorithm are used to identify trends in the patient population.

Several classification algorithms are investigated for predictive analysis, like C & RT-DT, CAD, CHAID, and so on. According to the results, the C & RT-DT algorithm predicts CAD with the highest accuracy, with an error rate of 0.074 for the whole dataset. Performance does, however, differ amongst clusters; CHAID works best in cluster 3, while C & RT-DT performs best in clusters 1 and 2.

Because dealing with a substantial number of features might lead to forecasts being unreliable, the study emphasizes the significance of feature selection. Many factors, including the features of heart disease, affect how accurate CAD prediction is.

Reference [5] explored multiple techniques based on machine learning with the objective to develop a classification algorithm that accurately predicts cardiovascular diseases. Strategies like K-fold Cross-Validation were used to judge the model's performance and generalizability. With a total of around 918 rows, the dataset was split into training and testing sets. Some of the significant algorithms were AdaBoost, Gradient Boost, Random Forest, k-nearest neighbor, Support Vector Machine, and Decision tree classifiers. The findings showed that while recall and F1 ratings were somewhat lower for RandomForest, Logistic Regression, SVM, and KNeighbors, accuracy and precision scores were comparable. The results demonstrated that accuracy and precision scores were similar for Random Forest, Logistic Regression, SVM, and KNeighbors, while recall and F1 ratings were somewhat lower. XGBoost & AdaBoost revealed intermediate accuracy and precision, however worse recall and F1 scores.   
With 93% accuracy for cardiac sickness and 88% accuracy for no disease, Catboost obtained a total accuracy of 91%. This is an important development in the categorization of cardiovascular diseases.

Reference[6] uses a framework where a machine learning ensemble model, combines many classification methods to increase the precision with which cardiovascular illnesses are identified. On several datasets, the ensemble model achieved high accuracy rates, outperforming seven benchmark methods.

Results from the research indicate that the proposed ensemble model produced reliable outcomes, with accuracy rates of 88.24% on the Cleveland dataset, 93.39% on the whole dataset, and 96.75% on the Mendeley dataset. The model consistently performed well across multiple datasets, indicating its reliability and potential for application in the healthcare sector.

Reference [7]

The research utilized an ensemble stacking classifier for the classification of cardiovascular conditions and diabetes by integrating meta-classifiers (Random Forest and SVM) with elementary classifiers (KNN, Naive Bayes, Linear Discriminant Analysis, and Decision Trees).

The results of the research demonstrated that the ensemble stacking classifier exceeded individual classifiers when it came to overall accuracy, recall, precision, and F-measure for both diabetes and heart diseases. The accuracy percentages of the proposed stacking model had been greater than that of the individual classifiers. For it to accomplish this, Random Forest was used as the meta-classifier along with the four elementary classifiers. For example, the accuracy rates of LDA, SVM, and Decision Tree for heart disease were 0.8426%, 0.8472%, and 0.8523%, respectively. In contrast, the proposed stacking model's accuracy was 0.9735% for diabetes disease.

Reference [8] developed an ensemble stacking classifier to classify diabetes and cardiovascular diseases using an approach that integrated meta-classifiers (Random Forest and SVM) with primary classifiers (KNN, Naive Bayes, Linear Discriminant Analysis, and Decision Trees). Among the databases utilized by the study were the Statlog and Cleveland heart disease datasets. Although the Cleveland dataset consisted of 14 attributes with 13 input values and 1 target value, the Statlog dataset had 270 instances with both clinical and non-clinical features.

Reference [9]

This paper describes the studies that employed machine learning methods to forecast heart problems. The University of California, Irvine repository provided the dataset, which comprised 68,975 patient records with 14 attributes. The study made use of a number of classifiers, including random forests, decision trees, and support vector machines. Missing data was addressed, normalization, quality assessment, transformation, encoding, dimensionality reduction, and feature scaling were among the preprocessing operations performed on the dataset.

The findings revealed that the advised hybrid model surpassed previous prediction models and achieved excellent accuracy rates. Support vector machines made up the models with the highest training accuracy, decision tree and random forest, which were 99.980% and 99.30% accurate, respectively. The results of the research show that the proposed classifiers had excellent accuracy rates, indicating the effectiveness of machine learning algorithms in the diagnosis and prediction of cardiovascular diseases.

Reference [10]

The Heart UCI dataset and the Cardiovascular Disease dataset were the two cardiovascular datasets used in the study. The clinical information from study participants is included in the datasets, which also show whether or not cardiovascular disease is present. The datasets were divided into training, validation, and test sets after being preprocessed to put the data into the proper format. While Auto-Sklearn was used to automatically generate and test classifiers on the same datasets, the graduate student manually trained and tested machine learning models on the datasets.

The study's findings showed that AutoML, more especially Auto-Sklearn, produced classification accuracies that were on par with or even higher than the graduate student's manually trained models. Better generalization was indicated by the AutoML models' much higher area under the ROC curve and area under the precision-recall curve. The study discovered that AutoML could create competitive machine learning models fast and with little assistance from humans.

Reference [11]

The study's dataset comprised biosignals obtained during walking from senior citizens 65 years of age or above. ECG and PPG data, which were separated into distinct intervals for analysis, were included in the bio-signals.

The study's outcomes showed that the suggested method could correctly forecast the prognostic signs of stroke illness in the elderly. The following was the reported accuracy of the predictive models:

- C4.5 Decision Tree: 91.08%

- RandomForest: 97.51%

- CNN-LSTM: 99.15%

Reference [12]

The research paper is built on a modular web-based software framework that enables co-learning and in-depth study of cardiovascular imaging data. It is made up of several major elements, including a data model, a semi-automated tool for efficient labeling, an interface for visual analytics, and features for radiomics and cardiovascular analysis extraction. The framework establishes a connection with the clinical infrastructure through DICOM network services in order to receive imaging data from PACS systems.

The results of the study indicate that when paired with visual analysis tools, the expert-in-the-loop approach can speed up algorithm training and enhance the interpretability of taught features. With corrections of the classifier input parameters and training dataset

**References:**

[1] S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in IEEE Access, vol. 7, pp. 81542-81554, 2019.

[2] Hassan, Ch. Anwar ul, Jawaid Iqbal, Rizwana Irfan, Saddam Hussain, Abeer D. Algarni, Syed Sabir Hussain Bukhari, Nazik Alturki, and Syed Sajid Ullah. 2022. "Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers" Sensors 22, no. 19: 7227.

[3] Jagtap, A., Malewadkar, P., Baswat, O., & Rambade, H. (2019). Heart disease prediction using machine learning. *International Journal of Research in Engineering, Science and Management*, *2*(2), 352-355.

[4] Shamsollahi, M., Badiee, A., & Ghazanfari, M. (2019). Using combined descriptive and predictive methods of data mining for coronary artery disease prediction: a case study approach. *Journal of AI and Data Mining*, *7*(1), 47-58.

[5] Baghdadi, N.A., Farghaly Abdelaliem, S.M., Malki, A. et al. Advanced machine learning techniques for cardiovascular disease early detection and diagnosis. J Big Data 10, 144 (2023).

[6] Bhanu, Prakash, Doppala., S., Bhattacharya., Midhunchakkaravarthy, Janarthanan., Nam-Sik, Baik. (2022). A Reliable Machine Intelligence Model for Accurate Identification of Cardiovascular Diseases Using Ensemble Techniques. Journal of Healthcare Engineering, 2022:1-13.

[7] Khan A, Khan A, Khan MM, Farid K, Alam MM, Su'ud MBM. Cardiovascular and Diabetes Diseases Classification Using Ensemble Stacking Classifiers with SVM as a Meta Classifier. Diagnostics (Basel). 2022 Oct 26;12(11):2595. doi: 10.3390/diagnostics12112595. PMID: 36359438; PMCID: PMC9689866.

[8] Nabila Kausar, & Dr. Hamid Ghous. (2020). A Novel Framework for Classification of Cardiovascular Disease Using Local Linear Embedding as Feature Selection Method. LC International Journal of STEM (ISSN: 2708-7123), 1(3), 17-31.

[9] Azhar, Imran., Areesa, Ashfaq., Muhammad, Adnan. (2022). Exploiting Machine Learning Models for Identification of Heart Diseases. 3(02):1-20.

[10] Padmanabhan M, Yuan P, Chada G, Nguyen HV. Physician-Friendly Machine Learning: A Case Study with Cardiovascular Disease Risk Prediction. J Clin Med. 2019 Jul 18;8(7):1050. doi: 10.3390/jcm8071050. PMID: 31323843; PMCID: PMC6678298.

[11] J. Yu, S. Park, S. -H. Kwon, K. -H. Cho and H. Lee, "AI-Based Stroke Disease Prediction System Using ECG and PPG Bio-Signals," in IEEE Access, vol. 10, pp. 43623-43638, 2022, doi: 10.1109/ACCESS.2022.3169284.

[12] G. Revathy, P. M. Priya, K. Senthilnathan, P. Mythili and S. V. Haridharani, "GUI based Heart using Disease Classification using Machine Learning," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 330-333.

[13] “Cardiovascular Diseases (Cvds).” World Health Organization, World Health Organization www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). Accessed 6 Apr. 2024.

[14] Kishor A., Jeberson W. Diagnosis of heart disease using internet of things and machine learning algorithms; Proceedings of the Second International Conference on Computing, Communications, and Cyber-Security; Ghaziabad, India. 3–4 October 2020; pp. 691–702.

[15] Gupta G., Adarsh U., Reddy N.S., Rao B.A. Journal of Physics: Conference Series. Volume 2161. IOP Publishing; Bristol, UK: 2022. Comparison of various machine learning approaches uses in heart ailments prediction; p. 012010.

[16] Dulhare, U. N. (2018). Prediction system for heart disease using naïve bayes and particle swarm optimization. Biomedical Research, 29 (12), 2646-2649.

[17] Gawali, M., & Shirwalkar, N. (2018). Heart disease prediction system using data mining techniques. International Journal of Pure and Applied mathematics, 120 (6), 499-506.

[18] Haq, A. U., Li, J.-P., Memon, M. H., Nazir, S., & Sun, R. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. Hindawi Mobile Information System. doi: 10.1155/2018/3860146.

[19] Hariharan, K., Vigneshwar, W. S., Sivaramakrishnan, N., & Subramaniyaswamy, V. (2018). A comparative study on heart disease analysis using classification techniques. International Journal of Pure and Applied Mathematics, 119 (12), 13357-13366. [15] Hussein, M. U. (2017, October 29). Physics and the Cardiovascular System.

[20] Akash, Rajak., Rabi, Narayan, Panda. (2023). Prediction of Heart Disease from the Multiple Features Available in the Database by Applying Classification Models of Machine Learning. doi: 10.1109/AISP57993.2023.10134906

[21] Polydoros, N., Kampaktsis., Maria, Emfietzoglou., Aamna, Al, Shehhi., Nikolina-Alexia, Fasoula., Constantinos, Bakogiannis., Dimitrios, Mouselimis., Anastasios, Tsarouchas., Vassilios, Vassilikos., Michael, Kallmayer., Hans-Henning, Eckstein., Leontios, J., Hadjileontiadis., Angelos, Karlas. (2023). Artificial intelligence in atherosclerotic disease: Applications and trends. Frontiers in Cardiovascular Medicine, doi: 10.3389/fcvm.2022.949454

[22] (2022). Artificial Intelligence in Cardiovascular Medicine: Current Insights and Future Prospects. Vascular Health and Risk Management, doi: 10.2147/vhrm.s279337

[23] Avvaru, R, V, Naga, Suneetha., T., Mahalngam. (2022). Cardiovascular Disease Prediction Using ML and DL Approaches. International Journal on Recent and Innovation Trends in Computing and Communication, doi: 10.17762/ijritcc.v10i10.5745

[24] Chayakrit, Krittanawong., Chayakrit, Krittanawong., Hafeez, Ul, Hassan, Virk., Sripal, Bangalore., Zhen, Wang., Zhen, Wang., Kipp, W., Johnson., Rachel, Pinotti., Hong, Ju, Zhang., Scott, Kaplin., Bharat, Narasimhan., Takeshi, Kitai., Usman, Baber., Jonathan, L., Halperin., W.H., Wilson, Tang. (2020). Machine learning prediction in cardiovascular diseases: a meta-analysis. Scientific Reports, doi: 10.1038/S41598-020-72685-1