**Literature Review:**

Numerous studies have been conducted utilising diverse machine learning and deep learning methodologies. Researchers used various data mining techniques, including association rules, classification, and clustering, to develop a model for predicting cardiovascular disease. In this section, we have evaluated some of the most important prior work on the prediction of cardiovascular illness utilising ensemble-based machine learning and deep learning techniques.

K. Polaraju et al. [1] utilised Multiple Regression Model and demonstrated that Multiple Linear Regression is suitable for predicting cardiovascular disease chance. The job is carried out with a training data set consisting of consisting of 3000 copies data records with thirteen distinct features.  The data set is separated into two sections: 70% of the data is used for training and 30% for evaluation and testing. It is evident from the findings of the work that the classification is accurate which shows that the Regression algorithm is more precise than others algorithms.

Shiva Kazempour Dehkordi1 & Hedieh Sajedi suggested a method for prediction based on the prescription utilizing the data mining method [2]. They devised a method named Skating to improve the precision of the system. Skating is a group technique, similar to Boosting and Bagging methods. They compared four classification algorithms, including Decision Tree Classifiers, Naïve Bayes (NB), K-Nearest Neighbours and, (KNN) and Skating on a distinct label. They demonstrated that the most precise estimate was the supplied classifier is arduous. This categorization algorithm yielded 73.17 percent accuracy. However, this is a relatively lower score compared to other categorization methods and algorithms. For instance, Jan et al. adopted an ensemble data mining strategy in 2018 utilising two benchmark datasets acquired from the UCI repository (Cleveland and Hungarian) in which the ensemble of five distinct classification algorithms, including RF, neural network, NB, and classification using neural networks was used. Utilizing regression analysis and support vector machines (SVM), [3].They discovered in the study that the algorithm with the poorest performance was regression approaches, but RF supplied an extremely high a precision of 98.136%.

In [4], Venkatalakshmi and Shivsankar devised a diagnostic paradigm for cardiac disease. The model provided was based on Naive Bayes and decision trees. UCI's heart disease dataset served as the basis for the trials. Wake tool was applied to extract valuable characteristics from the dataset. The accuracy of the suggested model was 85.03% for Naive Bayes and 84.03% for decision tree. Maio et al. [5] created a predictive model of hospital mortality for patients with heart failure using an enhanced random survival forest. The tests utilised for the experiment was a public dataset from the MIMIC II clinical database consisting of 8059 patients and 32 characteristics. The proposed system's achieved the accuracy of 82.01%.

A ML model utilising SVM and Naive Bayes was proposed by Laskshmi and Haritha [6]. An online dataset from the Cleveland heart disease dataset was used in this study for the sake of the experiments. The ROC chart was used to validate the suggested model's outcome, and the reported accuracy was 84.87%. For better cardiac disease identification, Javeed et al. [7] presented an intelligent learning system based on a random search algorithm and optimised random forest model. The suggested diagnostic system uses the grid search algorithm for optimization and the random search technique for feature selection. The Cleveland dataset, an online database of heart failure patients, was used in experiments. Only 7 features total were included in the proposed approach to identify heart disease. The newly proposed technique achieved an accuracy of 93.33%.

Different data mining techniques were developed by Boshra Brahmi et al. ways to assess heart disease prediction and diagnosis disease. The primary goal is to assess the various Various categorization methods, including J48, Decision Tree, KNN, SMO and simplistic Bayes. After that, assessing a few performance on tests of precision, sensitivity, and accuracy evaluation and comparison of specificity and gave the average accuracy of 86%. The paper concluded that the decision tree and J48 is the most effective method for predicting heart disease for the chosen dataset [10].

Jyoti Soni et al. utilised Decision Trees with a genetic algorithm in 2011 to increase the performance of categorization, and this was assessed with other two techniques, including Naive Bayes and cluster-based classification algorithms [8]. The proposed approach demonstrated 99.2% accuracy. Similarity, Hend Mansoor et al. examined the performance of Logistic Regression and Random Forest classification algorithms for assessing CVD patient risk exposure [9]. They demonstrated that the LR Model performed better than the Random Forest Classification method. Logistic Regression Model had an accuracy of 89%, whilst RF Model had an accuracy of 88% accuracy. Conventional Logistic Regression model performed in the mentioned studies performed admirably in predicting the potential occurrence of CVD.

**References:**

[1] Chen, Irene Y., Emma Pierson, Sherri Rose, Shalmali Joshi, Kadija Ferryman, and Marzyeh Ghassemi. "Ethical machine learning in healthcare." *Annual review of biomedical data science* 4 (2021): 123-144.

[2] Qayyum, Adnan, Junaid Qadir, Muhammad Bilal, and Ala Al-Fuqaha. "Secure and robust machine learning for healthcare: A survey." *IEEE Reviews in Biomedical Engineering* 14 (2020): 156-180.

[3] Kim, Su Yeon, James Diggans, Dan Pankratz, Jing Huang, Moraima Pagan, Nicole Sindy, Ed Tom et al. "Classification of usual interstitial pneumonia in patients with interstitial lung disease: assessment of a machine learning approach using high-dimensional transcriptional data." *The lancet Respiratory medicine* 3, no. 6 (2015): 473-482.

[4] Austin, Peter C., Jack V. Tu, Jennifer E. Ho, Daniel Levy, and Douglas S. Lee. "Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes." *Journal of clinical epidemiology* 66, no. 4 (2013): 398-407.

[5] Dietterich, Thomas G. "Ensemble methods in machine learning." In *International workshop on multiple classifier systems*, pp. 1-15. Springer, Berlin, Heidelberg, 2000.

[6] Al’Aref, Subhi J., Khalil Anchouche, Gurpreet Singh, Piotr J. Slomka, Kranthi K. Kolli, Amit Kumar, Mohit Pandey et al. "Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging." *European heart journal* 40, no. 24 (2019): 1975-1986.

[7] K. Polaraju, D. Durga Prasad, “Prediction of Heart Disease using Multiple Linear Regression Model”, International Journal of Engineering Development and Research Development, ISSN: 2321-9939, 2017.

[8] S.K. Dehkordi, H. Sajedi, Prediction of disease based on prescription using data mining methods, Health Technol. 9 (1) (2018) 37–44.

[9] M. Jan, A.A. Awan, M.S. Khalid, S. Nisar, Ensemble approach for developing a smart heart disease prediction system using classification algorithms, Res. Rep. Clin. Cardiol. 9 (2018) 33–45.

[10]. B. Venkatalakshmi and M. Shivsankar, “Heart disease diagnosis using predictive data mining, international journal of innovative research in science,” *Engineering and Technology*, vol. 3, no. 3, pp. 1873–1877, 2014.

[11] F. Miao, Y.-P. Cai, Y.-X. Zhang, X.-M. Fan, and Y. Li, “Predictive modeling of hospital mortality for patients with heart failure by using an improved random survival forest,” *IEEE Access*, vol. 6, pp. 7244–7253, 2018.

[12] M. S. Lakshmi, D. Haritha, and V. SRKIT, “Heart disease diagnosis using predictive data mining,” *International Journal of Computer Science and Information Security*, 2016.

[13] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R. Nour, “An intelligent learning system based on random search algorithm and optimized random forest model for improved heart disease detection,” *IEEE Access*, vol. 7, pp. 180235–180243, 2019.

[14] J. Soni, U. Ansari, D. Sharma, S. Soni, Predictive data mining for medical diagnosis: an overview of heart disease prediction, Int. J. Comput. Appl. 17 (8) (2011) 43–48.

[15] H.M. Islam, Y. Elgendy, R. Segal, A.A. Bavry, J. Bian, Risk prediction model for inhospital mortality in women with ST-elevation myocardial infarction: a machine learning approach, J. Heart & Lung (2017) 1–7.

[16] Boshra Brahmi, Mirsaeid Hosseini Shirvani, “Prediction and Diagnosis of Heart Disease by Data Mining Techniques”, Journals of Multidisciplinary Engineering Science and Technology, vol.2, 2 February 2015, pp.164- 168.

[17] S. Palaniappan and R. Awang, “Intelligent heart disease prediction system using data mining techniques,” in IEEE/ACS International Conference on Computer Systems and Applications. IEEE, 2008, pp. 108– 115.

[18] Ackerman, Samuel, Eitan Farchi, Orna Raz, Marcel Zalmanovici, and Parijat Dube. "Detection of data drift and outliers affecting machine learning model performance over time." *arXiv preprint arXiv:2012.09258* (2020).

[19] Cai, Jie, Jiawei Luo, Shulin Wang, and Sheng Yang. "Feature selection in machine learning: A new perspective." *Neurocomputing* 300 (2018): 70-79.

[20] Benesty, Jacob, Jingdong Chen, and Yiteng Huang. "On the importance of the Pearson correlation coefficient in noise reduction." *IEEE Transactions on Audio, Speech, and Language Processing* 16, no. 4 (2008): 757-765.

[50] A. Ul Haq, J. Li, Z. Ali, J. Khan, M. H. Memon, M. Abbas, and S. Nazir, ‘‘Recognition of the Parkinson’s disease using a hybrid feature selection approach,’’ J. Intell. Fuzzy Syst., vol. 39, pp. 1–21, May 2020, doi: 10.3233/JIFS-200075.

[51] A. Ul Haq, J. Li, Z. Ali, J. Khan, M. H. Memon, M. Abbas, and S. Nazir, ‘‘Recognition of the Parkinson’s disease using a hybrid feature selection approach,’’ J. Intell. Fuzzy Syst., vol. 39, pp. 1–21, May 2020, doi: 10.3233/JIFS-200075.