Nat.Volatiles&Essent.Oils,2021;8(4):4341-4353

NVEO
Natural Volatiles & Essential Oils

Impact of Deep Learning Algorithms in Cardiovascular Disease Prediction

Bhavya S, Ranjana P

School of Computing Sciences, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu, India

Abstract

Heart disease is a life-threatening condition that can strike anyone anywhere in the world. If it can be predicted, then actions can be taken to avoid it. According to the World Health Organization (WHO), cardiovascular diseases (CVDs) are the largest cause of death worldwide, killing an estimated 17.9 million of people each year. In healthcare, a predictive model learns from a patient's past data in order to forecast their future illnesses and prescribe treatment. A deep learning model can help health professionals make quick judgments about medications and hospitalizations, saving time and improving the healthcare business. The use of deep learning and machine learning models like Support Vector Machine (SVM), Naive Bayes (NB), Random Forest (RF), Decision Trees (DT), Multilayer Perceptron Neural Network and K-nearest neighbor(KNN), Artificial Neural Network (ANN), Auto Encoder (AE) and Recurrent Neural Network (RNN) on various healthcare applications is highlighted in this review. The result indicates that, machine learning\deep learning model with feature selection and optimization techniques improves the prediction accuracy and effectiveness of the model by reducing the system's learning time. This paper analyzes the various studies conducted in cardiovascular disease prediction (CVD)using deep learning\machine learning approaches.

Keywords: Deep Learning (DL), Machine Learning (ML), cardiovascular disease (CVD), Prediction Models, ANN, Autoencoder, Electronic Health Records (EHRs)

1. Introduction

According to the World Health Organization (WHO), cardiovascular diseases (CVDs) are the largest cause of death worldwide, killing an estimated 17.9 million of people each year. Coronary heart disease, cerebrovascular illness, rheumatic heart disease, and other ailments are among the heart and blood vessel disorders known as CVDs. Heart attacks and strokes are responsible for more than four out of every five CVD deaths, with one-third of these deaths occurring before the age of 70. Unhealthy food, physical inactivity, cigarette use, and problematic alcohol consumption are the most major behavioural risk factors for heart disease and stroke. Individuals may experience symptoms such as high blood pressure, high blood glucose, high blood lipids, and overweight or obesity as a result of behavioural risk factors. These "intermediate risk factors" are detectable in primary care settings and signal a higher risk of heart attack, stroke, heart failure, and other consequences. Identifying people who are most at risk for CVDs and ensuring that they receive adequate treatment can help to reduce early death [1].

A deep learning model can assist healthcare professionals to make decisions regarding medications, hospitalizations quickly and thus save time and also serve the healthcare industry better. In healthcare, a deep learning-based predictive model is notably useful for detecting diseases early, predicting treatment, and recommending future hospitalizations. It is identified from literature review that deep learning/machine learning algorithms are capable enough for the early identification of onset of heart disease using patients' electronic health records (EHRs). This study provides an overview of the deep learning/ machine learning methodologies and algorithms that are used for cardiovascular disease (CVD) prediction.

2. Need for Deep Learning\Machine Learning in Disease Prediction

The main difference between statistical and DL\ML methods is that the former is generally used to study the relationships between a small number of variables, whereas the ML\DL method is used for identification and engineering of features from data as well as prediction. ML\DL methods thus supplement and extend conventional statistical methods by providing tools and algorithms for deciphering patterns from huge, complex, and heterogeneous data sets. ML\DL approaches are adaptable and generalizable over a number of data sources and enable analyses and interpretation across complicated variables. Furthermore, ML\DL approaches use fewer assumptions and produce superior and more reliable predictions [2-3].

3. Deep Learning\Machine Learning based CVD prediction models

A combination framework of support vector machines (SVM), logistic regression (LR), and decision trees (DT) accurately predicted the early prognosis of cardiovascular diseases. The authors' proposed a rule-based approach to assess the accuracies of applying rules to individual findings of support vector machines, decision trees, and logistic regression on the Cleveland Heart Disease Database in order to present an accurate model of heart disease prediction [4]. Also, the authors' concluded that more complicated and combinational models such as neural networks and ensemble algorithms are needed to improve the accuracy of forecasting the early beginning of cardiovascular diseases.

A heart disease diagnosis system was proposed by using a rough sets-based attribute reduction and interval type-2 fuzzy logic system (IT2FLS). IT2FLS employs a hybrid learning approach that includes a fuzzy c-mean clustering algorithm as well as chaos firefly and genetic hybrid algorithms for parameter tuning [5]. Heart Disease Dataset (catalog) and SPECTF dataset from UCI Repository Machine Learning Database were used for the study. First, binary particle swarm optimization and rough sets-based attribute reduction are compared to the chaos firefly algorithm and rough sets-based attribute reduction. Second, three well-known classification techniques, namely Naive Bayes (NB), Support Vector Machine (SVM), and Artificial Neural Network (ANN), are compared to the type-2 fuzzy logic system (proposed approach). According to results, the proposed model received an accuracy of 88.3% for heart disease dataset and for SPECTF dataset, the model received an accuracy of 87.2% which were higher than the other classifiers [5].

In [6], A Recurrent Neural Network (RNN) model with Gated Recurrent Unit (GRU) was developed to predict the Heart Failure cases by using EHR of patients from Sutter Palo Alto Medical Foundation. With a 12- to 18-month observation window of patients and controls, this network could find relationships between time-stamped events (e.g., disease diagnosis, drug orders, procedure orders, etc.). According to results, the area under the curve (AUC) for the RNN model was 0.777 using a 12-month observation window, compared to AUCs for logistic regression (0.747), multilayer perceptron (MLP) with one hidden layer (0.765), support vector machine (SVM) (0.743), and K-nearest neighbor (KNN)(0.730). The AUC for the RNN model climbed to 0.883 while employing an 18-month observation window, which was significantly higher than the 0.834 AUC for the best of the baseline approaches (MLP). According to study, with a shorter observation window of 12–18 months, deep learning models tuned to leverage temporal relations appear to increase model performance for identification of incident heart failure.

A hybrid neural network- genetic algorithm accurately predicted the diagnosis of coronary artery disease. The proposed method improved the performance of a neural network by 10% by increasing its initial weights with a genetic algorithm that proposed better weights for the neural network. The model had an accuracy of 93.85% on Z-Alizadeh Sani dataset [7]. The authors used a feed forward neural network structure with one hidden layer with 5 neurons, 22 input features and an output layer that predicted whether the patient had heart disease or not.

The authors used Cerner Health Facts® EMR (Electronic Medical Record) data, to test an RNN (Recurrent Neural Network) model called RETAIN (REverse Time AttentIoN) for heart failure onset risk prediction [8]. Over 150,000 heart failure patients and over 1,000,000 controls were included in the study, which came from over 400 hospitals. In comparison to logistic regression, RETAIN attained an AUC of 82%, illustrating the power of more expressive deep learning models for EHR predictive modelling. The accuracy of the predictions varied by patient category and by hospital. In addition, the authors tested RETAIN models on specific hospitals and discovered that the model may be applied to other hospitals with only a 3.6% loss in AUC.

For the diagnosis of cardiac disease, a hybrid intelligent machine-learning-based prediction system was proposed using Cleveland heart disease dataset [9]. Seven classifiers such as LR (logistic regression), K-NN (K-Nearest Neighbor), ANN (Artificial Neural Network), SVM (Support Vector Machine), NB (Naive Bayes), DT (Decision Tree), and RF (Random Forest)were used, along with three feature selection techniques, Relief, mRMR(minimal redundancy-maximal-relevance), and LASSO (Shrinkage and Selection Operator). The system was validated using the K-fold cross-validation method. When the feature selection algorithm Relief chose the classifiers logistic regression with 10-fold cross-validation, the best accuracy was 89%. It is a more accurate predictive method in terms of accuracies because to the superior performance of logistic regression with Relief. When compared to the specificity of logistic regression with feature selection methods Relief and LASSO, SVM (linear) with a feature selection, algorithm mRMR performed the best in terms of specificity- 100%.ANN (MLP) with 16 hidden neurons and with Relief feature selection method has the highest sensitivity on selected characteristics, at 100%. The sensitivity of the Naive Bayes classifier with the LASSO FS method is the lowest (78%). Also, the results proved that the use of feature selection algorithms to get the optimal features, improved the classification accuracy and lowered the diagnosis system's execution time [9].

In [8], the authors proposed a hybrid model, HRFLM, by combining the benefits of both Random Forest (RF) and Linear Method (LM) to predict the onset of heart disease by using the Cleveland Heart

Disease Dataset. The model was compared with other methods such as Naive Bayes, Generalized Linear Model, Logistic Regression, Decision Tree, Random Forest, Gradient Boosted Tree and Support Vector Machine and received a better accuracy of 88.4% than other classifiers. The authors suggested that novel feature selection approaches may be developed to gain a broader view of important aspects and improve heart disease prediction performance [10].

The accuracy of heart disease prediction using an ensemble of classifiers is investigated in another research. For training and testing, the Cleveland heart dataset from the UCI machine learning repository was used. For the experiments, the ensemble algorithms bagging, boosting, stacking, and majority voting were used. The accuracy was enhanced by a maximum of 6.92% when bagging was applied. The accuracy was enhanced by a maximum of 5.94% when boosting was utilised. The accuracy of the weak classifiers was improved by a maximum of 7.26% when they were ensembled with majority voting, and by a maximum of 6.93% when they were stacked. Results proved that using feature selection approaches, the performance was improved even more. The feature selection methods improved the accuracy of the ensemble algorithms [11].

The researchers developed a method for evaluating whether a patient has heart disease based on cardiac sounds alone. The sounds of the patient's heartbeat were recorded using a smartphone and the iStetho scope Pro mobile app as the study's dataset [12]. The authors gathered a set of features derived directly from cardiac sounds and used this feature vector as input to a deep neural network with four hidden layers and an output layer to determine if a cardiac sound belongs to a healthy or a diseased patient. In the first three hidden layers, the relu activation function was utilised, and in the last hidden layer, softmax was used. The proposed method has a precision of 0.98, demonstrating its use in discriminating between healthy and ill patients [12].

A heart disease prediction system was proposed, which covered steps such as feature extraction, record and attribute minimization, and classification. The first step was to extract both statistics and higher-order statistical characteristics, followed by record and attribute minimization.PCA (Principal Component Analysis) was used to solve the "curse of dimensionality". These dimensionally reduced features were supplied to NN (Neural Network), which then performed the prediction procedure. Because the proposed study is primarily concerned with accurate prediction, it was decided to optimise the weights of NN. The authors used the PM-LU (Particle Swarm Optimization Merged Lion Algorithm Update) model for optimization, which combines the concepts of LA (Lion Algorithm) and PSO (Particle Swarm Optimization). Finally, the proposed model’s performance was compared to that of other traditional methods. The study was on the Heart Disease dataset (statlog) which was downloaded from UCI Machine learning repository. According to the results, the proposed PM-LU-NN method was 3.85% more accurate than LM-NN (Levenberg–Marquardt- NN), 12.5% more accurate than WOA-NN (Whale Optimization Algorithm- NN), 12.5% more accurate than FF-NN (Firefly- NN) 3.85% more accurate than PSO-NN, and 7.41% more accurate than LA-NN algorithms respectively in terms of accuracy [13].

The authors proposed a model that combined a chi-square (CHI) with Principal Component Analysis (PCA) to improve machine learning model prediction. The classifier's purpose was to predict whether or not a patient had heart disease. According to authors, when system resources must be considered, using all features is not possible. The study effectively investigated the use of dimensionality reduction techniques to improve the raw data outcomes. The authors chose three groupings of attributes from the 74 available and attained the best results. The Heart Disease dataset obtained from UCI Machine Learning Repository used for the study. CHI-PCA with RF (Random Forest) had the best performance with 98.7% accuracy for Cleveland, 99.0% accuracy for Hungarian, and 99.4% accuracy for CH (Combination of Cleaveland and Hungarian dataset) among the other classifiers such as Decision Tree, Gradient-Boosted Tree, Logistic Regression, Multilayer Perceptron, Naive Bayes and Random Forest [14].

In another study, authors proposed a deep learning approach, Universal Sentence Encoder (USE) to detect the duplicate records in the dataset and proved that the duplicate record elimination will increase the quality of classifier results. Cleaveland Heart Disease dataset was used for the study. To obtain the cleaned datasets versions, the authors used the duplicate records elimination method for each dataset. The C5.0 classification algorithm was then used to classify the records (or samples) in order to predict whether or not the patients had cardiac disease. Experiments showed that when the duplicate records eradication strategy is used to datasets with duplicate records, classification performance increases [15].

Different machine learning methods such as Neural Network (NN), Decision Tree (DT), K-Nearest Neighbor(KNN), Combined Nomenclature(CN2) rue inducer, Support Vector Machine (SVM), and Stochastic Gradient Descent (SGD) were implemented for the prediction of heart disease. The Heart Disease dataset (Statlog) downloaded from the 'UCI Machine Learning Repository'. The DT and SVM algorithms provided the greatest results on the 20-fold Cross-validation and 10-fold Cross-validation (87.69%) respectively. SGD also claims to have the best 5-fold Cross-validation scores (87.69%) [16].

An improved sparse auto encoder (SAE)-based ANN is proposed in [17] for heart disease prediction. The sparse auto encoder was used to learn the optimal representation of the data, and the ANN was used to make predictions based on the records learned. The Adam method was used to optimise the SAE, and batch normalisation was used. The model had a 90% accuracy rate on test data. The authors claimed that the proposed strategy performed better than certain traditional machine learning algorithms such as KNN (81% accuracy), Classification and Regression Tree (CART- 76% accuracy), Logistic Regression (83%accuracy), Naive Bayes (82%accuracy), Linear Discriminant Analysis (LD- 83% accuracy) and ANN (85% accuracy). The results proved that sparse auto encoder learning of low-dimensional features improved the ANN's classification performance [17].

The authors used heart disease dataset from Kaggle, to investigate the usefulness of machine learning approaches to heart disease prediction and discovered that three classification algorithms, KNN, RF, and DT, performed exceptionally well with 100% accuracy. In addition, for all of the applied algorithms except MLP and KNN, feature importance ratings for each feature were computed. The feature importance score was used to rank these features. The goal of this research was to find the most accurate machine learning classifiers for diagnostic purposes [18].

A heart disease prediction framework was proposed by using main risk variables and classifier algorithms including Naive Bayes (NB), Bayesian Optimized Support Vector Machine (BO-SVM), K-Nearest Neighbor (KNN), and Salp Swarm Optimized Neural Network (SSA-NN).The heart disease dataset provided on the UCI Machine Repository is used in this study to aid in the accurate detection of heart disease. The best results were obtained with BO-SVM with an accuracy of 93.3% than other classifiers such as KNN with accuracy of 80%, NB with 86.7%, NN with 80%, SSA-NN with 86.7%, SVM with 80% of accuracy respectively [19]. The results proved that the proposed unique optimised algorithm can be used to create a reliable healthcare monitoring system for the early detection of heart disease.

In [20], the authors compared the various classifications and CVD forecasts. Data Mining Techniques for CVD, Machine Learning Models for CVD, and Deep Learning Models for CVD prediction were organised in a three-part structure. According to the review results, Decision Tree and Naive Bayes Classifiers are the most extensively used classification methods. According to the results, Neural Network-based classification outperforms the other two approaches.

Other data mining methods, such as Clustering, Association Rule based, and Time Series based, canon the other hand, be examined for the use of predictions. According to machine learning aspects, the general conclusions drawn from the implemented models are: More medical attributes can be used to create a better model that is more accurate and performs better. Also, the combination of data-mining and text-mining techniques with existing models can be implemented to create effective prediction systems. Neural network-based solutions should be improved to meet a wide range of heart-related disorders, according to observations from deep learning approaches. In the case of ANN, we can alter the design and train algorithms to achieve more precise results [20].

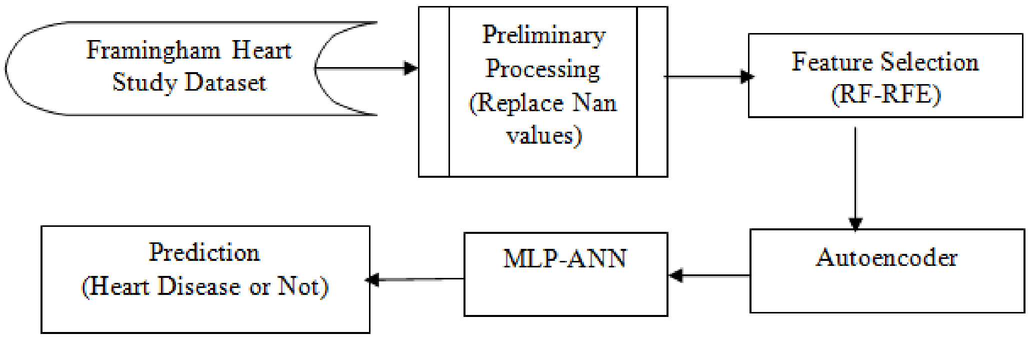
4. A proposed deep learning model to predict CVD

An improved deep learning-based CVD prediction model will be proposed by incorporating all the suggestions and recommendations from the literature review.

The Heart Disease (HD) dataset is downloaded from Kaggle website. It was a cardiovascular cohort study of residents from the city of Framingham, Massachusetts and it has a higher number of instances (4238) compared to other HD datasets such as Cleveland, Hungarian and Long Beach Heart datasets which have 303, 294, and 200 samples respectively.

An Autoencoder- Artificial Neural Network (AE-ANN) model will be implemented. Initially, the feature selection algorithms will be applied and the reduced feature set will be given as an input to the autoencoder model. The low-dimensional features learned by autoencoder model will be feed as an input to the ANN and accuracy will be evaluated. Researchers already proved that the combination of AE-ANN has improved the performance of model accuracy than a stand-alone ANN [15]. To build an efficient deep learning model with better accuracy, the proposed method will be implemented with the feature selection technique. The following diagram (Fig1.) shows the architecture of the proposed work.

Figure 1. Workflow of Proposed Method



5. Conclusion

Various deep learning\machine learning models used in CVD prediction are reviewed in this paper. From the literature review it is identified that, feature selection strategies play a critical role in the classification\prediction process. The more important attributes aid in the accurate classification of medical data, which is necessary for disease diagnosis. As a result, feature selection is crucial since it can exclude illness signs that aren't important. It also improves the effectiveness of the model by reducing the system's learning time. Analyzing the extensive EHR information of patients is time-consuming, and deep learning can help doctors and healthcare professionals discover and predict diseases based on scans, MRI reports, EHR, and other sources. It is observed from the literature review that deep learning models with effective feature selection strategies enhanced cardiovascular disease prediction accuracy. The proposed model will be implemented by using an efficient deep learning model, AE-ANN with feature selection algorithm to get the better accuracy.

Following Table 1 depicts the summary of the related work mentioned in the paper with the result observed.

Table 1. An overview of related work in terms of Methods used, Sample data and the Results obtained from each.

Related Work	**Methods Used**	**Sample Data**	**Results**
Mythili, T et al. (2013)- A heart disease prediction model using SVM-Decision Trees-Logistic Regression (SDL)	Rule based approach on SVM, LR and DT	Cleveland Heart Disease Database from UCI- ML Repository	Authors' Observation: more complicated and combinational models such as neural networks and ensemble algorithms are needed to improve the accuracy of forecasting the early beginning of CVDs.
Long, N.C. et al. (2015)- A highly accurate firefly-based algorithm for heart disease prediction	Interval type-2 fuzzy logic system (IT2FLS) with fuzzy c-mean clustering; chaos firefly	Heart Disease Dataset (catalog) and SPECTF from UCI- ML Repository	Accuracy of: 88.3% for heart disease dataset;and 87.2% for SPECTF dataset, which

	and genetic hybrid algorithms for parameter tuning		were higher than the other classifiers.
Choi, E. et al. (2017)- Using recurrent neural network models for early detection of heart failure onset	Recurrent Neural Network (RNN) model with Gated Recurrent Unit (GRU)	EHR of patients from Sutter Palo Alto Medical Foundation	The AUC (Area Under the Curve) for the RNN model was 0.883 for 18-month observation window, which was significantly higher than the 0.834 AUC for the best of the baseline approaches (MLP).
Arabasadi, Z et al. (2017)- Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm	Hybrid neural network-genetic algorithm	Z-Alizadeh Sani dataset	Accuracy of the model: 93.85%
Rasmy, L et al. (2018)- A study of generalizability of recurrent neural network-based predictive models for heart failure onset risk using a large and heterogeneous EHR data set	RNN called, RETAIN (REverse Time AttentIoN) model	Cerner Health Facts® EMR data	AUC of model, RETAIN- 82%. Also, the authors tested RETAIN model on specific hospitals and discovered that the model may be applied to other hospitals with only a 3.6% loss in AUC.
Haq, A.U. et al. (2018)- A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms	Used seven classifiers- LR, K-NN, ANN SVM, NB, DT and RF with three feature selection algorithms- Relief, mRMR and LASSO.	Cleveland Heart Disease dataset	The more accurate method in terms of accuracy was LR with Relief- 89% of accuracy. Also, results proved that the use of an efficient feature selection algorithm improves accuracy of the model.
Mohan, et al. (2019)- Effective heart disease prediction using hybrid machine learning techniques.	Hybrid model- HRFLM (Random Forest (RF) and Linear Method (LM))	Cleveland Heart Disease dataset	The model received a better accuracy of 88.4% than other classifiers such as NB, Generalized Linear Model, LR, DT, RF, Gradient Boosted Tree and SVM. As per authors' novel feature

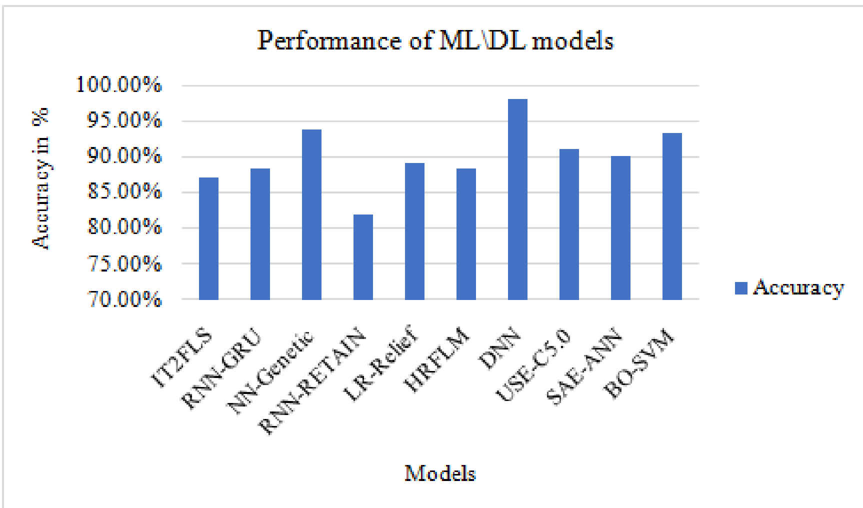
			selection approaches may be developed to gain a broader view of important aspects and improve heart disease prediction performance
Latha and Jeeva. (2019)- Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques	Ensemble of classifiers- bagging, boosting stacking, and majority voting	Cleveland Heart Disease dataset	Bagging- The accuracy was enhanced by 6.92%; Boosting- accuracy improved by 5.94%; Majority voting- accuracy improved by 7.26% and for stacking accuracy improved by 6.93%. Also, the feature selection methods improved the accuracy of the ensemble algorithms.
Brunese, L et al. (2019)- Deep learning for heart disease detection through cardiac sounds	Deep Neural Network	Patients' Cardiac Sounds were recorded using a smartphone and the iStethoscope Pro mobile app	The proposed method has a precision of 0.98.
Cherian, R.P. et al. (2020)- Weight optimized neural network for heart disease prediction using hybrid lion plus particle swarm algorithm	Neural Network with Particle Swarm Optimization Merged Lion Algorithm Update (PM-LU-NN)	Heart Disease dataset (statlog)	The proposed PM-LU-NN method was 3.85% more accurate than LM-NN (Levenberg–Marquardt- NN), 12.5% more accurate than WOA-NN (Whale Optimization Algorithm-NN), 12.5% more accurate than FF-NN (Firefly- NN) 3.85% more accurate than PSO-NN, and 7.41% more accurate than LA-NN algorithm
Gárate-Escamila et al. (2020)- Classification models for heart disease prediction using feature selection and PCA	chi-square with Principal Component Analysis (CHI-PCA)	Heart Disease Dataset (Cleaveland and Hungarian)	The model received an accuracy of 98.7% for Cleveland, 99.0% accuracy for Hungarian, and 99.4% accuracy for CH (Combination of Cleaveland and

			Hungarian dataset). Also, the study proved the use of dimensionality reduction techniques to improve the accuracy of model.
Lattar, H et al. (2020)- Does data cleaning improve heart disease prediction?	Universal Sentence Encoder (USE) with C5.0 classification algorithm	Cleveland Heart Disease dataset	According to authors, when the duplicate records eradication strategy is used to datasets with duplicate records, classification performance increases; for 10% of duplicate records, the model received 91% of accuracy.
Pires, I.M. et al. (2020)- Machine learning for the evaluation of the presence of heart disease	Used six ML models: NN, DT, KNN, CN2, SVM and SGD	Heart Disease dataset (Statlog)	Accuracy obtained: DT with 20-fold cross validation- 87.69%; SVM with 10-fold cross validation- 87.69%; SGD with 5-flod cross validation- 87.69%
Mienye, I.D. et al. (2020)- Improved sparse autoencoder based artificial neural network approach for prediction of heart disease	Sparse Autoencoder (SAE)-based ANN	Framingham Heart Study Cohort	SAE-ANN received 90% accuracy than a stand-alone ANN (85%)
Ali, M.M et al. (2021)- Heart disease prediction using supervised machine learning algorithms	ML methods such as RF, KNN, DT and MLP	Heart Disease dataset from Kaggle	KNN, RF, and DT, performed exceptionally well with 100% accuracy
Nayak, G.S et al. (2021)- Heart disease prediction by using novel optimization algorithm: A supervised learning prospective.	NB, Bayesian Optimized SVM (BO-SVM), KNN, and Salp Swarm Optimized Neural Network (SSA-NN)	Heart Disease dataset from UCL ML Repository	BO-SVM received higher accuracy of 93.3% than other classifiers such as KNN with accuracy of 80%, NB with 86.7%, NN with 80%, SSA-NN with 86.7%, SVM with 80% of accuracy respectively

The table summarizes that majority of the authors have chosen the two popular heart disease datasets such as Cleaveland Heart disease dataset and Statlog Heart disease dataset from UCL machine learning repository for their study to predict CVD. Machine/deep learning algorithms such as SVM (Support Vector Machine), Recurrent Convolutional Neural Network (RNN), Naive Bayes (NB), K-nearest neighbor (KNN), Decision Trees (DT), Artificial Neural Network (ANN) etc. were used widely for the studies. It is also observed that the deep learning algorithms: ANN, Autoencoder and MLP with different feature selection techniques gave the better accuracy on a large volume of dataset. Also, can observe that use of feature selection algorithms to get the optimal features, improved the prediction accuracy and lowered the diagnosis system's execution time. By considering all these factors, a proposed deep learning approach (AE-ANN) will be implemented to improve the accuracy of CVD prediction model.

The following graph depicts the accuracy obtained for each method mentioned in the literature review:

Figure 2. Performance of ML\DL Models in terms of accuracy



References

Cardiovascular diseases (who.int)

Shameer, K., Johnson, K.W., Glicksberg, B.S., Dudley, J.T. and Sengupta, P.P., 2018. Machine learning in cardiovascular medicine: are we there yet?. Heart, 104(14), pp.1156-1164.

Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science 2015;349:255–60

Mythili, T., Mukherji, D., Padalia, N. and Naidu, A., 2013. A heart disease prediction model using SVM-Decision Trees-Logistic Regression (SDL). International Journal of Computer Applications, 68(16).

Long, N.C., Meesad, P. and Unger, H., 2015. A highly accurate firefly based algorithm for heart disease prediction. Expert Systems with Applications, 42(21), pp.8221-8231.

Choi, E., Schuetz, A., Stewart, W.F. and Sun, J., 2017. Using recurrent neural network models for early detection of heart failure onset. Journal of the American Medical Informatics Association, 24(2), pp.361-370.

Arabasadi, Z., Alizadehsani, R., Roshanzamir, M., Moosaei, H. and Yarifard, A.A., 2017. Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. Computer methods and programs in biomedicine, 141, pp.19-26.

Rasmy, L., Wu, Y., Wang, N., Geng, X., Zheng, W.J., Wang, F., Wu, H., Xu, H. and Zhi, D., 2018. A study of generalizability of recurrent neural network-based predictive models for heart failure onset risk using a large and heterogeneous EHR data set. Journal of biomedical informatics, 84, pp.11-16.

Haq, A.U., Li, J.P., Memon, M.H., Nazir, S. and Sun, R., 2018. A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. Mobile Information Systems, 2018.

Mohan, S., Thirumalai, C. and Srivastava, G., 2019. Effective heart disease prediction using hybrid machine learning techniques. IEEE access, 7, pp.81542-81554.

Latha, C.B.C. and Jeeva, S.C., 2019. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Informatics in Medicine Unlocked, 16, p.100203.

Brunese, L., Martinelli, F., Mercaldo, F. and Santone, A., 2020. Deep learning for heart disease detection through cardiac sounds. Procedia Computer Science, 176, pp.2202-2211.

Cherian, R.P., Thomas, N. and Venkitachalam, S., 2020. Weight optimized neural network for heart disease prediction using hybrid lion plus particle swarm algorithm. Journal of Biomedical Informatics, 110, p.103543.

Gárate-Escamila, A.K., El Hassani, A.H. and Andrès, E., 2020. Classification models for heart disease prediction using feature selection and PCA. Informatics in Medicine Unlocked, 19, p.100330.

Lattar, H., Salem, A.B. and Ghezala, H.H.B., 2020. Does data cleaning improve heart disease prediction? Procedia Computer Science, 176, pp.1131-1140.

Pires, I.M., Marques, G., Garcia, N.M. and Ponciano, V., 2020. Machine learning for the evaluation of the presence of heart disease. Procedia Computer Science, 177, pp.432-437.

Mienye, I.D., Sun, Y. and Wang, Z., 2020. Improved sparse autoencoder based artificial neural network approach for prediction of heart disease. Informatics in Medicine Unlocked, 18, p.100307.

Ali, M.M., Paul, B.K., Ahmed, K., Bui, F.M., Quinn, J.M. and Moni, M.A., 2021. Heart disease prediction using supervised machine learning algorithms: performance analysis and comparison. Computers in Biology and Medicine, 136, p.104672.

Nayak, G.S., Patro, S.P. and Padhy, N., 2021. Heart disease prediction by using novel optimization algorithm: A supervised learning prospective.

Swathy, M. and Saruladha, K., 2021. A comparative study of classification and prediction of Cardio-vascular diseases (CVD) using Machine Learning and Deep Learning techniques. ICT Express.

Murtaza, Mustafa, M. Kamarudin, and M. Phanindranathan. "Clinical effectiveness of c-reactive protein as inflammation marker in cardiovascular disease." International Journal of General Medicine and Pharmacy (IJGMP) 2 (2013): 37-4.

Padmanaban, S., and Martin L. William. "A nonparametric discriminant variable-selection algorithm for classification to two populations." International Journal of Applied Mathematics and Statistical Sciences 5.2 (2016): 87-98.

Sankar, A., S. R. Gunakala, and D. Comissiong. "The effect of heat transfer on two-layered blood flow through a composite stenosis in the presence of a magnetic field." International Journal of Applied Mathematics and Statistical Sciences 3 (2014): 17-28.

Radhika, Rani L., and S. Chandralingam. "Acoustic Analysis of Cry Signal To Differentiate Healthy and Congenital Heart Disorder in Infants." International Journal of Electronics and Communication Engineering (IJECE) 5.6, Oct - Nov 2016; 21-28

Abbas, Hamed Jadooa, and Jawad Mohammad Ismail. "Association of Adiponectin Levels in Iraqi Patients With Type 2 Diabetic and Prediction of Cardiovascular Complications." Tjprc: International Journal of Diabetes & Research (Tjprc: Ijdr) 2.1 (2016) 1 8 (2016).

Shakeel, Mohd, Jamil Mohammad, and Mohammad Khalid Farooqui. "Study of C-Reactive Protein and Fibrinogen in Hypertension and Their Role in Development of Hypertension." International Journal of Applied, Physical and Bio-Chemistry Research (IJAPBCR) 4.4 (2014): 1-6.