

A Review of Machine Learning Techniques Used in the Prediction of Heart Disease

Chilkaragi Shankar Chaithra*^{ID}, Shivarudraswamy Siddesha^{ID}, V.N. Manjunath Aradhya^{ID},
Shanmukharadhya Keragodu Niranjana^{ID}

Department of Computer Applications, JSS Science and Technology University, Mysuru 570006, India

Corresponding Author Email: chaithracs@jssstuniv.in

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ria.380120>

ABSTRACT

Received: 31 August 2023

Revised: 20 October 2023

Accepted: 30 October 2023

Available online: 29 February 2024

Keywords:

heart disease prediction, classical machine learning, transfer learning, reinforcement learning, cardiovascular diseases

Heart disease stands as a principal cause of death worldwide, and its early prediction is essential for effective patient management and the reduction of healthcare expenditures. In this context, machine learning (ML) has emerged as a transformative tool in the healthcare sector, demonstrating a profound capability to discern intricate data patterns and furnish accurate prognostic assessments. The application of ML in cardiology is instrumental for risk prediction, early detection, and the customization of treatment protocols. The current study systematically reviews the spectrum of ML approaches applied to the prediction of heart disease, spanning supervised, unsupervised, reinforcement, and transfer learning methodologies. Data from prominent repositories such as Kaggle and the UCI Machine Learning Repository were employed to evaluate the performance of various ML algorithms, with key metrics including accuracy, sensitivity, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Influential predictors, namely age, gender, cholesterol levels, blood pressure, and lifestyle factors, were integral to the development of these predictive models. Particular attention was given to the exploration of ensemble methods and deep learning frameworks, which have shown to augment prediction accuracy beyond that of traditional models. This research delineates essential risk factors associated with heart disease and underscores the significance of predictive analytics in the healthcare landscape. With a focus on heterogeneous datasets and analytical techniques, the review aims to inform public health strategies and contribute to the alleviation of healthcare burdens. The elucidated findings highlight the promise of ML, particularly through the utilization of ensemble and deep learning methods, in the precursory prediction of heart disease. Such advancements enable healthcare professionals to make more informed decisions, adopt preventative interventions, and mitigate the overall impact on healthcare systems. This exhaustive review also synthesizes the efficacy and practicality of various ML algorithms, providing a valuable compendium for future research initiatives and promoting the integration of cutting-edge technologies in the management of cardiac health.

1. INTRODUCTION

Heart disease has vigorously increased over the years throughout the globe. Identifying and diagnosing the disease in the early stage is important for control death rate. Heart disease like cardiovascular disease, congenital heart disease, heart attacks, strokes, heart failures, blood vessel blockage etc. are the main cause for 31% of the death rate according to WHO [1]. Cardiovascular disease remains the highest cause of mortality in both men and women globally. In 2023 world population is around 8 billion, around 620 million people are heart patients all over the world. Around 60 million people develop heart disease each year. 1 in 3 deaths is because of heart diseases, an estimated 20.5 million deaths in 2021 that is one death in every 1.5 second says British heart foundation. [2]. The clinical services should be affordable to patients and avoid the burdens of unnecessary diagnosis, it requires proper diagnosis and accurate measure to be taken to provide effective treatment. Early-stage prediction and diagnosis

reduce the cost of treatment and heart disease mortality [3]. Causes of heart disease include high blood pressure which damages the arteries which causes accumulation of plaque. High cholesterol is also one of the leading causes for heart disease which builds up the plaque in arteries and restricts the blood flow to the heart. Smoking damages the blood vessels and causes plaque buildup which increases heart disease. Diabetes also increases the risk of heart disease in which high blood sugar level damages the arteries over time. Obesity and physical inactivity which contributes to obesity, high BP also heart disease risk factors. Poor diet which rich in trans fats, cholesterol, and sodium also the reason for development of heart disease. Excessive alcohol consumption increases blood pressure which causes heart failure [4-11].

Early prediction is very challenging to cardiologists due to many factors such as asymptomatic cardiovascular disease which is difficult to diagnose in early stage in which patient may don't recognize, limited diagnostic tools like cholesterol and blood pressure tests are always not sensitive or specific to

detect in early signs of CVD's hence doctors rely on imaging techniques like ECG etc. which may not routinely performed without clear symptoms, genetic and environment factors such as diet, lifestyle, exposure to polluted environment also plays major role in increasing CVD, lack of awareness about heart disease symptoms and risk factors leads to delayed medical diagnosis [12-15]. Heart disease is increasing due to behavioral risk factors like lifestyle of an individual, food intake, lack of activities, non-healthy habits also genetic factors, aging, environmental changes and environmental changes due to urbanization. Many individuals are affected in their younger age by heart problems. With this concern an automated machine learning system can be used to predict the disease in early stages. Machine learning algorithms learn from past clinical data and predict and gives an accurate result to reduce the burden of health practitioner [16]. Machine learning algorithms are complex procedures which identify the patterns and help in decision making by learning from trained datasets. Many Machine learning algorithms such as supervised learning which learns from data labels, unsupervised learning which learns without data labels, transfer learning which learns from tasks and reused to increase the performance and reinforcement learning which learns from trial and error has its efficient way to predict the heart disease [17].

Machine learning had played vital role in health care significantly in heart diseases such as risk prediction by analyzing the individual patient data, ML models helps in early detection of disease by analyzing the patterns of medical data, ML algorithms assist in analyzing image data such as MRI, CT etc. ML can be integrated to decision support system, ML algorithms also used to personalize treatments plans for individual patient, ML enabled wearable devices helps in continuous monitoring of patient's vital signs and irregularities which helps in reducing hospital readmissions, ML algorithms also used in drug discovery process to identify the potential compounds for heart disease treatment. Machine learning techniques such as data collection, feature selection and preprocessing are the initial way to provide the input to the machine learning algorithms then, efficient algorithm selection is also important for processing like supervised, ensemble or deep learning. Evaluation metrics are used to measure the efficiency of algorithms. Many research studies proved the efficiency of algorithms in prediction of heart disease [18-19].

There are many studies made to discover which type of machine learning technique gives more accurate results in diagnosing heart disease [20]. The proposed work includes a detailed review of different types of machine learning techniques for prediction of heart disease and different parameters to be considered for accurate prediction.

2. RELATED WORK

2.1 Classical learning (Supervised)

In supervised learning algorithms training dataset includes labelled data and non-labeled data set is used to test the algorithm for accurate result [3].

A model [16] is trained with 1025 Kaggle dataset which is a combination of four different datasets (Cleveland, Hungary, Long beach V and Switzerland) for heart disease prediction. Out of six classifiers Support Vector Machine (SVM),

Decision Tree (DT), K-nearest Neighbor (KNN), Naive bayes (NB), Random Forest (RF), DT achieved more accuracy about 98%. A model [1] was built for survival prediction of heart failure patients using DT, Decision tree Regressor (DTR), RF, XG Boost (XGB) and Gradient Boost (GB) machine learning algorithms. All algorithms are compared with respect to accuracy, precision, recall, F-measure, and log loss, RF gave highest accuracy of 97.78%. In one of the research work NB, SVM and DT classifiers used with 10 cross validations on 462 instances of south African heart dataset, NB performed well with accuracy of 71.5% [21]. A study proposed a study which uses Multilayer Perceptron (MLP), AdaBoost M1 (ABM1), KNN, DT, Logistic Regression (LR) and RF algorithm to predict the heart disease on Kaggle dataset in which RF achieved 100% accuracy in all precision, f-measure, and recall.[3]. A novel approach [22] was introduced by combining Neural Network (NN) for training and DT for test classification for prediction of heart disease on two integrated datasets (Cleveland and Stalog heart disease dataset), the model achieved 99.9 ROC area. An exploratory study [23] explains that model is trained on 1670 dataset from territory hospital south India and evaluated for five different classifiers (KNN, NB, LR, Ada Boost and RF), Random Forest algorithm achieved 93.8% accuracy.

A non-invasive method [24] used for predicting heart disease using LR and Stochastic Gradient Decent on 303 patient's record from UPI repository in which LR achieved accuracy of 91.67% and SDG with 80.0%. Classification methods of machine learning and image fusion are better in providing good accuracy for predicting heart diseases [25]. Classification and regression tree (CART) model achieved 87% accuracy in prediction of heart disease by extracting decision rules [26]. Feature selection technique called minimum redundancy and maximum relevance (MRMR) was applied after data preprocessing and discretization on UCI dataset, algorithms such as DT, LR, LG-SVM, NB and RF were used for prediction in which SVM achieved 84.85% accuracy [27]. Analysis on ML classifier algorithms [28] such as LG, KNN and RF is done on five combinations of datasets (Cleveland, Hungarian, Stalog, Switzerland and Long Beach datasets) out of three classifiers both RF and LG achieved 88% accuracy.

A performance evaluation was performed [29] on supervised ML algorithms (LR, DT, RF, SVM, NB, KNN) for heart disease prediction on Kaggle dataset, Boruta algorithm is used to select essential features after preprocessing, RF achieved highest accuracy of 83.53%. Hyperparameter tuning was made for reducing features for logistic regression by using randomized CV and grid CV method. Seven different combinations of algorithms were used to compare results. Combination of LR and Grid- Search with ensemble technique of Kernel PCA achieved 100% accuracy in which power transformation applied for preprocessing and Extra tree classifier is used for feature selection [30].

Heart disease prediction was done using logistic regression on UCI dataset, model achieved good accuracy about 87.10% for 90: 10 training and testing ratio [31]. Combination of ML and deep learning method was proposed [32] for predicting heart disease on four datasets (Cleveland, Hungary, Switzerland, and Long Beach V dataset). Data preprocessing is done by using Isolation Forest for detection of outliers, Lasso algorithm used for feature selection for which model achieved 94.2% accuracy.

A system was built for effective heart disease prediction

using NN in which multilayer perceptron neural network (MLPNN) with back propagation (BP) is used to train the model using 303 records of dataset and the system achieved 100% accuracy [33]. Different machine learning algorithms like ANN, RF and SVM were applied on 303 records from UCI repository for predicting the heart disease in which SVM performed well with accuracy of 84.0% [34]. An ensemble network was applied on 3 different datasets, algorithms like LR, KNN, RF, EXT, XGBoost and Ensemble classifiers were compared for predicting the heart disease in which XGBoost achieved 91% of accuracy [35]. A framework using deep learning based ensemble method was constructed for heart

disease prediction on Cleveland dataset from UCI repository, data preprocessing is done using K-bins discretization, Elliptic envelope, Randomized search CV, Isolation Forest and reduced model component analysis. The framework compares the result with SVC, AdaBoost, MLP, KNN etc., where GB performed well accuracy of 0.98 [36]. An attempt was made using ML and deep learning algorithms for predicting risk of coronary heart disease on Cleveland dataset. Model trained using SVM, KNN, DT, RF and ANN and compared their result after data preprocessing which ANN achieved higher accuracy [37]. The summary of the detailed discussion on the above can be seen in Table 1.

Table 1. Summary of classical learning (supervised) techniques

Authors	Research Focus	Technique(s) Applied	Dataset	Results	Limitations / Future Work
Ali et al. [1]	Risk factor analysis and survival prediction of heart failure patients	5 ML algorithms are used. DT, RF, GB, DTR, XG-Boost	Kaggle 299 instances	DTR: 85.24% XGB: 94.44% DT: 96.77% GB: 96.77% RF: 97.78%	
Ali et al. [3]	Identify ML classifiers of highest accuracy for HD prediction performance analysis	ADBM1, MLP, KNN, DT, RF	Kaggle 1025 patients record	LR: 89.62% ABM1: 95.02% MLP: 97.95% KNN: 100% DT: 100% RF: 100% SVM: 68.83% DT: 98.05% KNN: 87.01% LR: 75.97% NB: 75% RF: 95.13% NB: 90.4 ROC	
Thushar et al. [16]	Prediction of heart disease using different classifiers	DT, SVM, KNN, LG, NB, RF	Kaggle 1025 dataset	SVM: 82.1 ROC VP: 68.6 ROC NN: 88.9 ROC DT: 88.9 ROC NNDT: 99.9 ROC	i. Dataset: limited in number.
Hassani et al. [22]	Novel approach for predicting HD using hybrid neural network and decision tree	NNDT (neural network and decision tree)	Cleveland and Stalog heart disease datasets from UCI repository 573 records	LR: 91.67% SGD: 80.0%	ii. ML techniques can be combined for better accuracy.
Miranda et al. [24]	Early prediction heart disease using logistic regression and stochastic gradient descent	LG and SGD	303 medical records from UPI	LR 90: 10=87.10% LR 80: 20=85.25% LR 70: 30=83.52% LR 60: 40=81.97% LR 50: 50=81.58% RF: 83.52% SVM: 82.42% KNN: 72.53% LR: 80.22% DT: 79.12%	iii. Real time dataset can be applied on model.
Ambrish et al. [31]	Prediction of heart disease using logistic regression	LR	303 records of UCI repository		
Sujatha et al. [29]	Prediction of heart disease using supervised ML algorithms	RF, SVM, KNN, LR, DT	303 records from Kaggle		
Ambesange et al. [30]	Heart disease prediction with ensemble and hyper parameter tuning techniques	Preprocessing- power transformation yeo Johnson Feature engineering=extra tree classifier/ correlation matrix Algorithms=LR+ Grid Search CV	303 records from UCI repository	Randomized Search CV+LR: 84.94% RSCV+GSCV: 87.1% LR: 90.32% LR+GSCV: 100%	

Singh et al. [33]	Effective heart disease prediction using neural networks	MLPNN with BP	303 medical records	MLPNN with BP: 100%
Rindhe et al. [34]	Heart prediction using ML algorithms	ANN, SVM, RF	303 medical records	SVM: 84.0% NN: 83.5% RF: 80.0% ANN: 84.4%
Gangadhar et al. [37]	Coronary heart disease prediction using deep learning and ML techniques	SVM, KNN, DT, RF, ANN	Cleveland dataset	SVM: 83.33% RF: 81.67% DT: 73.33% KNN: 61.67% KNN: 88% NB: 88.2% LR: 90.8% AB: 92.6% RF: 93.8%
Maini et al. [23]	Heart disease prediction for south Indian population	KNN, NB, LR, AB, RF	1670 medical records of territory hospital	
Ozcan et al. [26]	Classification and regression tree is used to extract decision rules from features to predict heart disease	Classification and Regression tree	745 health records from heart disease dataset comprehensive, IEEE data port	CART: 87%
Bashir et al. [27]	Prediction of heart disease using feature selection approaches	DT, LG, LG SVM, RF, NB MRMR/FS	UCI repository	DT: 82.22% LR: 82.56% RF: 84.17% NB: 84.24% LG SVM: 84.85%
Patidar et al. [28]	Comparative analysis of ml algorithms for heart disease prediction	KNN, LG, RF	UCI dataset 1190 records	KNN: 73% LR: 88% RF: 88%
Bharti et al. [32]	Heart disease prediction using ML and deep learning approach	Preprocessing- Isolation Forest Feature selection=Lasso algorithm Classifiers=LR, K neighbors, SVM, RF, DT Deep learning	Cleveland, Hungary, Switzerland, and Long Beach V dataset	LR: 83.3% K neighbors: 84.8% SVM: 83.2% RF: 80.3% DT: 82.3% DL: 94.2
Nirmala et al. [35]	Prediction of heart disease using ensemble network and artificial intelligence	LR, KNN, RF, ET, XG boost, Ensemble classifier	Stat log, Hungary, Cleveland	KNN: 81% RF: 90% ET: 90% XGB: 91% Ensemble: 89% AB=0.87 SVM=0.92 RF=0.910 MLP=0.97 GB=0.98 KNN=0.89 DT=1.0 Gaussian NB=0.79
Venkatesh et al. [36]	Heart disease prediction using cloud computing, edge computing and ensemble techniques	AdaBoost, SVM, RF, MLP, GB, KNN, DT, Gaussian NB	Cleveland dataset	

2.2 Classical learning (Unsupervised)

Unsupervised learning is a form of machine learning in which algorithms analyze and cluster the datasets which are unlabeled. This learning will identify the hidden patterns without human help. Publications on clustering techniques have been done for prediction of heart disease.

A system was developed using classification and clustering for predicting heart dataset in which proposed model used K-means clustering where functions are clustered using Euclidian formula and normalization is applied for preprocessing the data. Logistic regression is used for classifying two attribute values of high cholesterol and borderline cholesterol. The model achieved accuracy of 70.58% without normalization and with normalization 84.84% and using classification 90% accuracy has achieved [38]. An integrated decision-making system was constructed for

predicting heart disease on Cleveland dataset. The system used principal component analysis (PCA) for dimensionality reduction and agglomerative for clustering and random forest for classification. The system is compared with various classifiers results for accuracy which is achieved was 95.65% which is higher than RF, DT, SVM, KNN, NB and LR classifiers [39].

A clustered particle swarm optimization technique used to train model on Cleveland dataset, the aim is to reduce the distance between training and testing data. K means clustering is applied to reduce inter cluster distance and particle swarm optimization technique to consider each cluster as separate optimization problem. The results were compared with base classifiers like DT, RF, SVM, KNN and NB where the proposed method performed well with accuracy of 96.03% [40]. An attempt was made using Principle Component Analysis (PCA) and Hybrid Genetic Algorithm (HGA) with k-

means to predict heart disease on 303 medical datasets from UPI repository. PCA is used to reduce attributes and hybrid k-means is used for clustering the data. The dataset is divided into PC1 and PC2 using PCA for main attributes. HGA is used to reduce fitness and improve the quality of clustering. Steady state algorithm is used to improve the clustering quality. The proposed method achieved accuracy of 94.06% [41].

In one of the studies [42], a clustered genetic optimization algorithm is used to predict heart disease on UCI dataset. Data segmentation is done using k-means algorithm and classification is done using genetic optimization technique and results are compared with base classifiers like DT, SVM, KNN, RF, NB etc., the proposed model achieved the accuracy of 94.56%.

Risk prediction of cardiovascular disease and diabetes using two stage semi supervised clustering method was built considering biomarkers. For this study US centers for disease

control and prevention (CDC) and National health and nutrition examination surveys (NHANES) data about 7508 for cardiovascular disease analysis and 5389 for diabetes were used after removing outliers. Semi supervised K-means clustering is used to divide the data into two groups that are low risk people and high-risk people who are about to face disease in future. Results are not disclosed in terms of accuracy in the paper [43]. An article expressed [44] an effective way of modeling the dataset for predicting heart disease using Particle swarm optimization (PSO) and A-priori algorithm. Proposed method used k-means clustering for grouping the data and PSO algorithm for rules and data behavior and relationships between the data and a hybrid k-means and PSO clustering is used for statistical analysis on parameters of data. The article only describes the modeling of medical data of heart for better understanding for prediction of heart disease.

Table 2. Summary of classical learning (Unsupervised) techniques

Authors	Research Focus	Technique(s) Applied	Dataset	Results	Limitations/ Future Work
Singh et al. [38]	Prediction of heart disease using clustering and classification	K-means clustering Logistic regression for classification Normalization (min -max)	Heart disease dataset	Clustering with normalization: 70.58% Clustering: 84.84% Classification: 90% LR: 89.13% NB: 84.78% SVM: 86.96% KNN: 65.22% DT: 91.30% RF: 93.48% IDMS: 95.65%	Different techniques can be incorporated to increase the accuracy of model.
Pati et al. [39]	Integrated decision-making system for predicting heart disease (IDMS)	Dimensionality reduction=PCA Clustering=Agglomerative clustering Classification=RF, LR, NB, SVM, DT, KNN	Cleveland dataset		Further investigations can be done on different datasets.
Islam et al. [41]	Early prediction of heart disease using PCA and HGA with k-means	Dimensionality reduction=PCA Clustering=HGA with k-means	UCI repository dataset	Clustering: 94.06%	Method can be used on different dataset.
Vijaya and Rao [40]	Heart disease prediction using clustered particle swarm optimization technique	Segmentation=K-means clustering Classification=clustered PSO	Cleveland dataset	Clustered PSO: 96.03% DT: 88.32% RF: 89.27% SVM: 89.53% KNN: 87.79% NB: 85.11% Clustered GA: 94.56%	Model can be tested for improved version of other optimization technique like whale and lion
Vijaya [42]	Clustered genetic optimization algorithm for heart disease prediction	Segmentation=Kmeans algorithm Classification=genetic optimization technique	303 records from UCI repository	DT: 88.49% RF: 89.48% SVM: 85.44% KNN: 87.16% NB: 88.52% NB: 52.33% DL: 52% KNN: 45.67%	
Ema and Shill [45]	Integrated model for heart disease prediction using fuzzy c means, artificial neural network and principal component analysis	clustering=FCM feature extraction=PCA	4200 records from UPI repository	[DATA MINING] NB: 86.53% DL: 89% ANN: 85.53% [IHDP] PCA_ANN: 85.35% FCM_ANN: 99.55%	PCA and FCM lacks in terms of precision, F-score and recall.

A model was designed by integrating Fuzzy c-means (FCM), PCA and ANN for predicting heart disease and trained with

4200 data from UPI repository, data preprocessing is done by calculating mean value for dataset, PCA is used to extract

important features and FCM is used to group the data into clusters. The model is built upon ANN architecture for both PCA and FCM and compared the results with various algorithms, the PCA combined with ANN achieved about 85.35% and FCM combined ANN achieved about 99.55% accuracy [45]. The above detailed discussion is tabulated briefly in Table 2.

2.3 Advanced learning

2.3.1 Transfer learning

Transfer learning is a type of machine learning where a model is trained for a particular task and is used for performing another task [46].

A method for predicting and classification of heart disease was developed using machine learning and transfer learning to select the proper features for training, validation to increase accuracy. Data preprocessing is done initially to reduce the dimensionality and to find the margin rate conjugant scaling factor is applied, to find the support values. Disease prone impact rate (DPIR) and to select the labelled data relative feature margin selection (RFMS) is used and model is trained with multilayer perception neural network (MLPNN). The model predicts the cardiac deficiency rate which has the higher impact. The model accuracy is compared with base classifiers for both prediction and accuracy. The purpose of this study is to predict heart disease towards accessing the big data which is analyzed in health care. The model achieved 96% accuracy

for 200 records for classification and 82% accuracy for prediction [47].

2.3.2 Reinforcement learning

Reinforcement learning is a form of machine learning in which the model learns from mistakes in the interactive environment [48].

A study combined reinforcement learning with multitask time series network to predict the grade of heart disease using color doppler echocardiograph, blood biochemical indicators and 10 parameters on boy information. The model performs in two steps in first pretraining is done using asynchronous advantage actor critic and trained data is adopted to RCNN for stochastic policy. In the second step soft and hard parameter sharing and time series network used to predict the heart disease [49] which achieved good accuracy about 0.9372 when compared to other models.

A model was developed for predicting heart disease using reinforcement learning (RL). The model is trained with Cleveland dataset of 303 records which uses Q-learning framework of RL for prediction. By developing the off-premised RL the model used only three parameters such as trestbps, chol and age and instructs the agent to determine the best rules for parameters. The accuracy of the proposed method is evaluated by using KNN and DT. The model achieved an accuracy of 0.8798 which is higher than KNN and DT [50]. A summary of the above discussion is shown in Table 3.

Table 3. Summary of advance learning techniques

Authors	Research Focus	Technique(s) Applied	Dataset	Results	Limitations/ Future Work
Sivaprasad et al. [47]	Heart disease prediction and classification using transfer learning models for predictive risk analysis of heart attack	RFMS-MLPNN	200 records	classification accuracy RF: 86% SVM: 90% HDPM: 92% RFMS-MLPNN: 96% Prediction accuracy RF: 76% SVM: 75% HDPM: 80% RFMS-MLPNN: 82%	Method can be applied on large number of datasets.
Prasanna et al. [50]	Heart disease prediction using Q-learning framework of reinforcement learning	Q-learning framework KNN, DT	Cleveland dataset of 303 attributes	Q-learning: 0.8798 DT: 0.7715 KNN: 0.7524	Only three attributes are considered to predict the disease. Work can be implemented for larger dataset and considering a greater number of parameters.

3. PERFORMANCE MEASURES

The different performance measures used for evaluating the reported models are accuracy, precision, recall, F-score and sensitivity and specificity. Performance of classification models is derived by confusion matrix. The confusion matrix is a table which the instances assigned to each class by which calculation can be done in terms of accuracy, AUC, true negative, true positive, false negative, false positive, sensitivity, specificity among other classifiers [21]. The evaluation performance of the above-mentioned model is

compared with other ML classifiers in terms of accuracy, precision, and recall. Tables 1, 2 and 3 describe the various learning techniques applied on dataset using base classifiers and advance techniques with their accuracy achieved. Random Forest classifier achieved good accuracy in prediction of heart disease [1, 3, 23, 28, 29]. In another study the model gave good accuracy, precision, and recall for Decision Tree [16]. In one of the studies KNN also achieved good accuracy in prediction [20]. In one of the model DT and NN outperformed in terms of accuracy, precision, recall, F-measure, sensitivity, specificity, and ROC area [21]. RF and SGD performed well

in prediction of heart disease [24]. In a few studies SVM performed well in terms of accuracy [27, 28, 34]. In the above-mentioned models, the combination of LG and grid search performed well [30] and in a few methods deep learning performed well [32, 36]. The ensembled methods achieved good accuracy when compared to base classifiers [35] and a combination of machine learning deep learning also gives good accuracy [37] and integrated system also achieved expected accuracy considering k-means clustering [38-42, 44].

Deep learning-based methods given promising results in risk prediction by combining classification and feature augmentation tasks and to extract new features from existing features sparse encoder and the convolutional classifier is combined in the proposed system which gave 90% accuracy when compared to the classical algorithms [51]. A novel method of deep learning is used to detect heart disease in which grey wolf horse herd optimization-based Shepard convolutional neural network algorithm (GWHHO-based ShCNN). The detection is done based on spark architecture which has slave node for preprocessing and feature fusion and master node for heart disease detection. Feature fusion is done using z-score normalization and missing value imputation. Feature fusion is performed using Hellinger distance with deep Q network which achieved 0.93 of accuracy [52]. The long short-term memory (LSTM) and recurrent neural network (RNN) based smart healthcare system achieved 99.9% accuracy in forecasting heart disease. Kalman filter is used to remove noisy data and to gather missing data, lion and krill optimization techniques were combined for feature extraction process [53].

Deep learning showed promising results as discussed in above section because of its ability to automatically learn patterns from large and complex datasets. It eliminates the need for manual feature engineering to extract the relevant features from raw data. Deep neural networks capture the complex relationships with hidden layers that traditional models. Deep learning algorithms require large amounts of data which is better for learning complex patterns. deep learning model can leverage pre trained networks on large dataset which can be transformed knowledge from one task to another using transfer learning when there is less labelled medical data. Deep learning models can be combined into ensembles to increase accuracy of algorithms. Advancement of CNN and for processing medical data and RNN for sequential data can improve the accuracy of heart disease prediction. Deep learning models' success depends on the quality and quantity of data provided to model. Deep learning algorithms can analyze various types of medical data like ECG signals, image data and medical records of patients to detect patterns which heart disease. Deep learning models can predict the risk factors and RNN can be used to analyze patient data and make predictions about future cardiac events. Deep learning models are used to personalize treatment by analyzing individual patient medical history and provides decisions about medications and it identifies genetic factors associated with heart disease [54-58].

4. DISCUSSION

4.1 Learning techniques

4.1.1 Supervised learning

Supervised ML algorithms consists of labeled datasets for training the model and testing on unlabeled datasets to

categorize them into similar groups. The supervised ML algorithms perform very well when datasets are sufficient and produce accurate results and predictions. Supervised ML algorithms required the efforts of humans to label the data and it takes long time to train the data. Supervised learning further classified into classification and regression models. Classification plays an important role in clinical research, in which supervised algorithms such as LR, KNN are easy to implement but the accuracy is reduced by noisy and irrelevant features. SVM is robust but doesn't handle noisy data, RF processes very large dataset for classification and regression, it can overcome missing values, but it is slow in producing predictions as it requires large dataset and trees. NN identifies the complexity of relationship between the variables, but it cannot access decision making process. NB is simple to implement handles complex data but there is loss of accuracy because it is based on assumption. DT is used widely to handle medical dataset for both numerical and categorical, the algorithm gives higher accuracy, but it may misclassify and only one attribute can be tested in decision making. Each ML algorithm has its own limitations. Each algorithm differs its results in different measures like accuracy, specificity, sensitivity, ROC, AUC, F-measure, recall and precision. Balancing the bi-variable response of a labeled class and increasing instances may result in improving ML algorithms performance. From the above study ensembled algorithms improve performance of ML based prediction system. For supervised ML algorithms input and output data is labelled and it is trained to recognize the relationship between input and output data. Feedback mechanism is used to validate predictions where number of classes are known [59-61].

4.1.2 Unsupervised learning

The goal of unsupervised learning is to classify the input values based on their similarities in which only input data is given. Unsupervised learning is used in clustering or grouping of similar objects, dimensionality reduction for reducing the dimensionality in datasets, and association for finding the relationship between variables. Unsupervised learning for predicting heart disease follows clustering techniques which groups similar and dissimilar types of data. Classification is done using classifiers for prediction analysis. K-means clustering algorithm is widely used for clustering, which is simple to implement, works on large dataset and it adapts for new dataset, but produces less accuracy when it clusters wrongly or some data remains un-clustered, and it may cluster the outliers. Hierarchical clustering is easy to implement which is divided into divisive and agglomerative clustering. Agglomerative clustering is used for grouping of objects based on similarity using bottom-up approach, it is good for small clusters. Divisive clustering is used to identify the large clusters. Unsupervised ML algorithms are used to analyze and cluster the unlabeled data where computing is complex, it does not provide feedback mechanism and works on real time data analysis [59, 62, 63].

4.1.3 Transfer learning

In this learning the knowledge is transferred from one model to another model where a model is trained for one task, and it is again reused for another task. Few studies have been done on heart disease prediction. Excess fit issues have been resolved by using transfer learning. Classical machine learning models can perform well on small datasets because models are very simple and very few parameters are considered. Deep NN

is complex, and data needs to be optimized for larger number of weights with larger training needed. Lot of manual work required for cleaning the data where millions of data points are needed and is time consuming. This also requires powerful resources for training. Transfer learning can be considered as a solution to these complex deep learning models, where it performs better with neural network, saves training time, and doesn't require lot of data [64].

Domain-adapted multichannel graph convolutional network (DAMGCN) was proposed as a novel method to transfer the knowledge between the graphs and cardiac datasets which outperformed existing single-domain network node classification methods to classify high-risk cardiac heart disease and low-risk cardiac heart disease [65]. Transfer learning is used to improve the trained convolutional neural network (CNN) to classify the heart rhythm from small ECG dataset. The model is pretrained with CNN on large dataset on continuous ECG dataset than it is finetuned on small dataset to classify Atrial Fibrillation which improved the result by 6.75% by reducing annotations which is required to achieve same performance as CNN without pretraining [66]. CNN based transfer learning is applied to predict lung disease using x-ray images, four models have been pre-trained to compare the results such as ResNet-50, MobileNet-V2, VGG-19 and DenseNet201 in which MobileNetV2 achieved about 98.45% accuracy [67]. A model is proposed using transfer and ensemble techniques to develop accurate, interpretable, and generalized models to predict heart disease. Transfer learning reduces the need for labeled data and ensemble method increase the robustness of model to noise and outliers in the data which achieved 90% accuracy [68].

4.1.4 Reinforcement learning

It is a type of learning based on feedback mechanism which uses leaning agent which makes observations and takes actions in the environment. It is used to achieve long term results and it corrects the errors while training the model, as model learns from own experience it doesn't require any data feeding. Reinforcement learning can be wrong when used as a framework as it requires a lot of data which can be expensive in case of computation. Reinforcement learning can be used with other techniques like deep learning [69, 70].

A model is proposed to identify weather conditions impacting the risk of congenital heart diseases. Reinforcement learning is used to build the model which accurately predict the weather conditions and identify interactions between air pollution levels and used to identify precipitation patterns which are associated with increase of congenital heart disease. The data collected related to weather and heart disease which includes temperature, humidity, air pressure, wind speed and used to identify the patterns of data which increase the risk of heart disease using algorithms. The model learns with trial-and-error methods taking the rewards by identifying the weather conditions which is associated with risk of increasing heart disease. Reinforcement learning requires a lot of data to train and to tune parameters, the accuracy depends on quality and quantity of data used. Measuring the results is more difficult in reinforcement learning because it uses trial and error method, and it requires large computational power such as expensive hardware [71].

4.2 Dataset

Dataset used for the heart disease prediction in the above

study were benchmark datasets from Kaggle and UPI repository namely Cleveland, Hungary, Switzerland, and Long Beach V dataset and a few of them were from hospitals. For analysis in the above studies parameters like age, sex, chest pain type, resting blood pressure, cholesterol, fasting blood sugar, ECG while resting, maximum heart rate, exercise induced angina, ST depression, slope of the peak, the number of major vessels colored and Thal is considered. About 14 to 16 attributes were considered from dataset as essential parameters to evaluate and analyze heart disease prediction. In one of the study ST slopes showed highest importance for predicting heart disease. Old peak, chest pain type, exercise angina also achieved maximum importance for prediction of heart disease cholesterol, resting ECG, resting blood pressure and age features achieved least importance when compared to other parameters. In a few of the research studies the datasets were integrated and preprocessed to remove outliers for accurate prediction. Since heart disease is based on the lifestyle of an individual, their food habits, BMI, hereditary, physical activity are most essential parts regardless of their blood pressure, diabetes, and other medical parameters. Most of the papers worked on available benchmark datasets are created and maintained for many years. The available dataset contains 14 to 16 parameters which are considered for heart disease prediction [26]. Many research studies are on prediction and detection of diseases [72, 73]. Some of the papers have made use of only three or four parameters for the study. This generation's lifestyle has been gradually changed over past years and we found that if we consider real-time data of recent medical records and a greater number of parameters of individual socio-economic activities, it would be better for accurate prediction of heart disease. Other than benchmark dataset parameters there could be other essential parameters, which can be considered for prediction of heart disease.

5. CONCLUSIONS

Heart disease has become a major risk factor worldwide due to which death rate is increased. Early detection is very important to minimize the rate of death and helps in automating the health sector which reduces the burden for cardiologists. This paper describes the detailed review of different learning techniques used in heart disease prediction help with higher accuracy. Research studies on classical and advance learning applied on heart disease prediction were selected from the past five years. Each learning technique applied in the above research studies gave good accuracy when it is used as integrated, data preprocessing is also very important before training the model. Dataset plays a major role when it comes to medical data, it is important to consider the socio-economic factors of an individual which leads to heart problems. According to World Heart Report 2023 risk factors for cardiovascular diseases include behavioral factors like insufficient physical activity, sodium intake, alcohol consumption, obesity, and smoking. Metabolic factors such as BP, fasting glucose, BMI, cholesterol, and diabetes. Environmental factors such as air pollution and heredity. In 2021 according to Global Burden of Disease Study modifiable risk factors such as tobacco use, high blood pressure contributed to 10.8 million CVD deaths globally. The risk factors are high and low depending upon geographical locations of the individual [74].

While collecting data, crucial parameters along with the

medical test results should be integrated to accurately predict the heart disease of an individual. Based on the above points there is still huge scope for timely prediction of heart disease in future. This study helps the researchers and students to know the limitations of the existing work and provides direction for future research. Future research required identifying the risk factors depending upon geographical area and individual lifestyle need furthermore study, work on reinforcement and transfer learning are fragmentary. Considering the number of parameters such as current lifestyle, eating habits and stress level of an individual might provide accurate result in prediction.

REFERENCES

- [1] Ali, M.M., Al-Doori, V.S., Mirzah, N., Hemu, A.A., Mahmud, I., Azam, S., Al-tabatabaie, K.F., Ahmed, K., Bui, F.M., Moni, M.A. (2023). A machine learning approach for risk factors analysis and survival prediction of Heart Failure patients. *Healthcare Analytics*, 3: 100182. <https://doi.org/10.1016/j.health.2023.100182>
- [2] World heart Day 2023: Reducing the burden of cardiovascular disease ... (n.d.). <https://www.pconline.com/News/Whats-new-on-PCRONline/2023/World-Heart-Day-2023-Reducing-burden-cardiovascular-disease-globally-beyond-stents-balloons>.
- [3] Ali, M.M., Paul, B.K., Ahmed, K., Bui, F.M., Quinn, J.M.W., Moni, M.A. (2021). Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. *Computers in Biology and Medicine*, 136: 104672. <https://doi.org/10.1016/j.combiomed.2021.104672>
- [4] Understanding Blood Pressure Readings. (n.d.). www.heart.org. <https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings>.
- [5] What is Blood Cholesterol? | NHLBI, NIH. (2022). NHLBI, NIH. <https://www.nhlbi.nih.gov/health-topics/high-blood-cholesterol>.
- [6] Health effects of smoking and tobacco use. (2022). Centers for Disease Control and Prevention. https://www.cdc.gov/tobacco/basic_information/health_effects/index.htm.
- [7] Diabetes complications | ADA. (n.d.). <https://www.diabetes.org/diabetes/complications>.
- [8] World Health Organization: WHO. (n.d.). Obesity and overweight. www.who.int. <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>
- [9] Current guidelines | Health.gov. (n.d.). <https://health.gov/our-work/physical-activity/current-guidelines>
- [10] The American Heart Association Diet and Lifestyle recommendations. (2023). www.heart.org. <https://www.heart.org/en/healthy-living/healthy-eating/eat-smart/nutrition-basics/aha-diet-and-lifestyle-recommendations>
- [11] Alcohol's effects on the body | National Institute on Alcohol Abuse and Alcoholism (NIAAA). (n.d.). <https://www.niaaa.nih.gov/alcohols-effects-health/alcohols-effects-body>
- [12] Lloyd-Jones, D.M., Hong, Y., Labarthe, D., Mozaffarian, D., Appel, L.J., Van Horn, L., et al. (2010). Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic Impact Goal through 2020 and beyond. *Circulation*, 121(4): 586-613. <https://doi.org/10.1161/CIRCULATIONAHA.109.192703>
- [13] Lopez, A.D., Mathers, C.D., Ezzati, M., Jamison, D.T., Murray, C.J. (2006). Global and regional burden of disease and risk factors, 2001: Systematic analysis of population health data. *The Lancet*, 367(9524): 1747-1757. [https://doi.org/10.1016/S0140-6736\(06\)68770-9](https://doi.org/10.1016/S0140-6736(06)68770-9)
- [14] Haffner, S.M., Lehto, S., Rönnemaa, T., Pyörälä, K., Laakso, M. (1998). Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. *New England Journal of Medicine*, 339(4): 229-234.
- [15] Update, A.S. (2017). Heart disease and stroke statistics—2017 update. *Circulation*, 135: e146-e603.
- [16] Tushar, A.M., Wazed, A., Shawon, E., Rahman, M., Hossen, M.I., Jesmeen, M.Z.H. (2022). A review of commonly used machine learning classifiers in heart disease prediction. In 2022 IEEE 10th Conference on Systems, Process & Control (ICSPC), pp. 319-323. <https://doi.org/10.1109/ICSPC55597.2022.10001742>
- [17] Khan, Y., Qamar, U., Yousaf, N., Khan, A. (2019). Machine learning techniques for heart disease datasets: A survey. In Proceedings of the 2019 11th International Conference on Machine Learning and Computing, pp. 27-35. <https://doi.org/10.1145/3318299.3318343>
- [18] Ghosh, P., Azam, S., Jonkman, M., Karim, A., Shamrat, F.J.M., Ignatious, E., Shultana, S., Beeravolu, A.R., De Boer, F. (2021). Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. *IEEE Access*, 9: 19304-19326. <https://doi.org/10.1109/ACCESS.2021.3053759>
- [19] Mohan, S., Thirumalai, C., Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*, 7: 81542-81554. <https://doi.org/10.1109/ACCESS.2019.2923707>
- [20] Shah, D., Patel, S., Bharti, S.K. (2020). Heart disease prediction using machine learning techniques. *SN Computer Science*, 1: 1-6. <https://doi.org/10.1007/s42979-020-00365-y>
- [21] Gonsalves, A.H., Thabtah, F., Mohammad, R.M.A., Singh, G. (2019). Prediction of coronary heart disease using machine learning: An experimental analysis. In Proceedings of the 2019 3rd International Conference on Deep Learning Technologies, pp. 51-56. <https://doi.org/10.1145/3342999.3343015>
- [22] Hassani, M.A., Tao, R., Kamyab, M., Mohammadi, M.H. (2020). An approach of predicting heart disease using a hybrid neural network and decision tree. In Proceedings of the 5th International Conference on Big Data and Computing, pp. 84-89. <https://doi.org/10.1145/3404687.3404704>
- [23] Maini, E., Venkateswarlu, B., Maini, B., Marwaha, D. (2021). Machine learning-based heart disease prediction system for Indian population: An exploratory study done in South India. *Medical Journal Armed Forces India*, 77(3): 302-311. <https://doi.org/10.1016/j.mjafi.2020.10.013>

- [24] Miranda, E., Bhatti, F.M., Aryuni, M., Bernando, C. (2021). Intelligent computational model for early heart disease prediction using logistic regression and stochastic gradient descent (A preliminary study). In 2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI), pp. 11-16. <https://doi.org/10.1109/ICCSAI53272.2021.9609724>
- [25] Diwakar, M., Tripathi, A., Joshi, K., Memoria, M., Singh, P. (2021). Latest trends on heart disease prediction using machine learning and image fusion. *Materials Today: Proceedings*, 37: 3213-3218. <https://doi.org/10.1016/j.matpr.2020.09.078>
- [26] Ozcan, M., Peker, S. (2023). A classification and regression tree algorithm for heart disease modeling and prediction. *Healthcare Analytics*, 3: 100130. <https://doi.org/10.1016/j.health.2022.100130>
- [27] Bashir, S., Khan, Z.S., Khan, F.H., Anjum, A., Bashir, K. (2019). Improving heart disease prediction using feature selection approaches. In 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), pp. 619-623. <https://doi.org/10.1109/IBCAST.2019.8667106>
- [28] Patidar, S., Jain, A., Gupta, A. (2022). Comparative analysis of machine learning algorithms for heart disease predictions. In 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1340-1344. <https://doi.org/10.1109/ICICCS53718.2022.9788408>
- [29] Sujatha, P., Mahalakshmi, K. (2020). Performance evaluation of supervised machine learning algorithms in prediction of heart disease. In 2020 IEEE International Conference for Innovation in Technology (INOCON), pp. 1-7. <https://doi.org/10.1109/INOCON50539.2020.9298354>
- [30] Ambesange, S., Vijayalaxmi, A., Sridevi, S., Yashoda, B. S. (2020). Multiple heart diseases prediction using logistic regression with ensemble and hyper parameter tuning techniques. In 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), pp. 827-832. <https://doi.org/10.1109/WorldS450073.2020.9210404>
- [31] Ambrish G., Ganesh, B., Ganesh, A., Srinivas, C., Mensinkal, K. (2022). Logistic regression technique for prediction of cardiovascular disease. *Global Transitions Proceedings*, 3(1): 127-130. <https://doi.org/10.1016/j.gltp.2022.04.008>
- [32] Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S., Singh, P. (2021). Prediction of heart disease using a combination of machine learning and deep learning. *Computational Intelligence and Neuroscience*, 2021: 8387680. <https://doi.org/10.1155/2021/8387680>
- [33] Singh, P., Singh, S., Pandi-Jain, G.S. (2018). Effective heart disease prediction system using data mining techniques. *International Journal of Nanomedicine*, 13(sup1): 121-124. <https://doi.org/10.2147/IJN.S124998>
- [34] Rindhe, B., Ahire, N., Patil, R., Gagare, S., Darade, M. (2021). Heart disease prediction using machine learning. *International Journal of Advanced Research in Science, Communication and Technology*, 267-276. <https://doi.org/10.48175/IJARSCT-11131>
- [35] Nirmala, S., Veena, K., Indu, B., Kalshetty, J.N. (2022). Heart disease prediction using artificial intelligence ensemble network. In 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), pp. 1-6. <https://doi.org/10.1109/MysuruCon55714.2022.9972493>
- [36] Venkatesh, V., Rai, P., Reddy, K.A., Praba, S., Anushiadevi, R. (2022). An intelligent framework for heart disease prediction deep learning-based ensemble Method. In 2022 International Conference on Computer, Power and Communications (ICCPC). pp. 274-280. <https://doi.org/10.1109/ICCPC55978.2022.10072285>
- [37] Gangadhar, M.S., Sai, K.V.S., Kumar, S.H.S., Kumar, K.A., Kavitha, M., Aravinth, S.S. (2023). Machine learning and deep learning techniques on accurate risk prediction of coronary heart disease. In 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), pp. 227-232. <https://doi.org/10.1109/ICCMC56507.2023.10083756>
- [38] Singh, R., Rajesh, E. (2019). Prediction of heart disease by clustering and classification techniques prediction of heart disease by clustering and classification techniques. *International Journal of Computer Sciences and Engineering*, 7(5): 861-866. <https://doi.org/10.26438/ijcse/v7i5.861866>
- [39] Pati, A., Parhi, M., Pattanayak, B.K. (2021). IDMS: an integrated decision making system for heart disease prediction. In 2021 1st Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology (ODICON), pp. 1-6. <https://doi.org/10.1109/ODICON50556.2021.9428958>
- [40] Vijaya, J., Rao, M. (2022). Heart disease prediction using clustered particle swarm optimization techniques. In 2022 IEEE 6th Conference on Information and Communication Technology (CICT), pp. 1-5. <https://doi.org/10.1109/CICT56698.2022.9997925>
- [41] Islam, M.T., Raza, S.R., Kibria, M.G. (2020). Early prediction of heart disease using PCA and hybrid genetic algorithm with k-means. In 2020 23rd International Conference on Computer and Information Technology (ICCIT), pp. 1-6. <https://doi.org/10.1109/ICCIT51783.2020.9392655>
- [42] Vijaya, J. (2023). Heart disease prediction using clustered genetic optimization algorithm. In 2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), pp. 1072-1077. <https://doi.org/10.1109/IITCEE57236.2023.10091050>
- [43] Mao, Z., Fukuma, Y., Tsukada, H., Wada, S. (2023). Risk prediction of chronic diseases with a two-stage semi-supervised clustering method. *Preventive Medicine Reports*, 32: 102129. <https://doi.org/10.1016/j.pmedr.2023.102129>
- [44] Goel, S., Singh, R. (2019). Modeling of heart data using PSO and A-Priori algorithm for disease prediction. In 2019 Fifth International Conference on Image Information Processing (ICIIP), pp. 475-479. <https://doi.org/10.1109/ICIIP47207.2019.8985732>
- [45] Ema, R.R., Shill, P.C. (2020). Integration of fuzzy C-means and artificial neural network with principle component analysis for heart disease prediction. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1-6. <https://doi.org/10.1109/ICCCNT49239.2020.9225366>
- [46] Brownlee, J. (2017). A gentle introduction to transfer learning for deep learning. *Machine Learning Mastery*, 20.

- [47] Sivaprasad, R., Hema, M., Ganar, B.N., Sunil, D.M., Mehta, V., Fahlevi, M. (2022). Heart disease prediction and classification using machine learning and transfer learning model. In 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), pp. 595-601. <https://doi.org/10.1109/ICACRS55517.2022.10029279>
- [48] Bhatt, S. (2019). Reinforcement learning 101 - towards data science. Medium. <https://towardsdatascience.com/reinforcement-learning-101-e24b50e1d292>.
- [49] Li, W., Zuo, M., Zhao, H., Xu, Q., Chen, D. (2022). Prediction of coronary heart disease based on combined reinforcement multitask progressive time-series networks. *Methods*, 198: 96-106. <https://doi.org/10.1016/j.ymeth.2021.12.009>
- [50] Prasanna, K.S., Challa, N.P., Nagaraju, J. (2023). Heart disease prediction using reinforcement learning technique. In 2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), pp. 1-5. <https://doi.org/10.1109/ICAECT57570.2023.10118232>
- [51] García-Ordás, M.T., Bayón-Gutiérrez, M., Benavides, C., Aveleira-Mata, J., Benítez-Andrades, J.A. (2023). Heart disease risk prediction using deep learning techniques with feature augmentation. *Multimedia Tools and Applications*, 82: 31759–31773. <https://doi.org/10.1007/s11042-023-14817-z>
- [52] Kanchanamala, P., Alphonse, A.S., Reddy, P.B. (2023). Heart disease prediction using hybrid optimization enabled deep learning network with spark architecture. *Biomedical Signal Processing and Control*, 84: 104707. <https://doi.org/10.1016/j.bspc.2023.104707>
- [53] Ramkumar, G., Seetha, J., Priyadarshini, R., Gopila, M., Saranya, G. (2023). IoT-based patient monitoring system for predicting heart disease using deep learning. *Measurement*, 218: 113235. <https://doi.org/10.1016/j.measurement.2023.113235>
- [54] Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1): 24-29. <https://doi.org/10.1038/s41591-018-0316-z>
- [55] Choi, E., Bahadori, M.T., Schuetz, A., Stewart, W.F., Sun, J.M. (2016). Doctor AI: Predicting clinical events via recurrent neural networks. *JMLR Workshop Conference Proceedings*, 56: 301-318.
- [56] Miotto, R., Wang, F., Wang, S., Jiang, X., Dudley, J.T. (2018). Deep learning for healthcare: Review, opportunities and challenges. *Briefings in Bioinformatics*, 19(6): 1236-1246. <https://doi.org/10.1093/bib/bbx044>
- [57] Min, S., Lee, B., Yoon, S. (2017). Deep learning in bioinformatics. *Brief Bioinform*, 18(5): 851-869. <https://doi.org/10.1093/bib/bbw068>
- [58] Rajkomar, A., Oren, E., Chen, K., Dai, A.M., Hajaj, N., Hardt, M., et al. (2018). Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine*, 1(1): 18. <https://doi.org/10.1038/s41746-018-0029-1>
- [59] Singh, R. (2020). A review on heart disease prediction using unsupervised and supervised learning. *Neural Networks*, 99: 100.
- [60] Uddin, S., Khan, A., Hossain, M.E., Moni, M.A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making*, 19(1): 1-16. <https://doi.org/10.1186/s12911-019-1004-8>
- [61] Jethva, H. (2023). Supervised learning vs unsupervised learning (Pros and cons). *Cloud Infrastructure Services*. <https://cloudinfrastructureservices.co.uk/supervised-learning-vs-unsupervised-learning/>.
- [62] K-means advantages and disadvantages |Machine Learning| Google for Developers. Google, developers.google.com/machine-learning/clustering/algorithm/advantages-disadvantages, accessed on 13 Aug. 2023.
- [63] How the Hierarchical Clustering Algorithm Works. *Dataaspirant*, 21 Dec. 2020, dataaspirant.com/hierarchical-clustering-algorithm/#-1608531820430.
- [64] Vinithavn. (2022). The power of transfer learning in deep learning - Analytics vidhya - medium. Medium. <https://medium.com/analytics-vidhya/the-power-of-transfer-learning-in-deep-learning-681f86a62f79>.
- [65] Lin, H., Chen, K., Xue, Y., Zhong, S., Chen, L., Ye, M. (2023). Coronary heart disease prediction method fusing domain-adaptive transfer learning with graph convolutional networks (GCN). *Scientific Reports*, 13(1): 14276. <https://doi.org/10.1038/s41598-023-33124-z>
- [66] Weimann, K., Conrad, T.O. (2021). Transfer learning for ECG classification. *Scientific Reports*, 11(1): 5251. <https://doi.org/10.1038/s41598-021-84374-8>
- [67] Mengistie, T.T., Kumar, D. (2021). Comparative study of transfer learning techniques for lung disease prediction. In 2021 10th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON), pp. 1-6. <https://doi.org/10.1109/IEMECON53809.2021.9689159>
- [68] Britto, C.F. (2023). Advancing heart disease prediction: integrating transfer and ensemble learning. In 2023 International Conference on Machine Learning in Health, Environment and Engineering Data.
- [69] Ramakrishnan, M. (2023) What is reinforcement learning's significance in AI development? *Emeritus Online Courses*. <https://emeritus.org/blog/ai-and-ml-what-is-reinforcement-learning/#:~:text=Reinforcement%20learning%20is%20a%20feedback,rules%20of%20the%20complex%20environment,> accessed on 15 August 2023.
- [70] Joy, A. (2022). Pros and cons of reinforcement learning. *Pythonista Planet*. <https://pythonistaplanet.com/pros-and-cons-of-reinforcement-learning/>.
- [71] Mohammed, M.A., Ramakrishnan, R., Mohammed, M.A., Mohammed, V.A., Logeshwaran, J. (2023). A novel predictive analysis to identify the weather impacts for congenital heart disease using reinforcement learning. In 2023 International Conference on Network, Multimedia and Information Technology (NMITCON), pp. 1-8. <https://doi.org/10.1109/NMITCON58196.2023.10276376>
- [72] R, C., Ashoka, D.V., B V, A.P. (2022). IMLAPC: Interfused machine learning approach for prediction of crops. *Revue d'Intelligence Artificielle*, 36(1): 169-174. <https://doi.org/10.18280/ria.360120>
- [73] Sarthi, G., Chikkaguddaiah, N., V. N., M.A. (2023). Human brain tumor detection and segmentation for MR

image. *Revue d'Intelligence Artificielle*, 37(1): 147-153.
<https://doi.org/10.18280/ria.370118>

[74] World heart report 2023 - world heart federation. (n.d.-

b). <https://world-heart-federation.org/wp-content/uploads/World-Heart-Report-2023.pdf>.