Amity Journal of Computational Sciences
ISSN 2456-6616 (Online)
Volume 7, Issue 1, 2023

Heart Disease Detection Using Machine Learning

Nikhil Kumar
Bachelors Of Technology
Greater Noida, Uttar Pradesh, India

I.ABSTRACT

Disease diagnosis is the most essential job in the healthcare industry. Early disease detection can save a lot of lives. The medical industry can greatly benefit from machine learning categorization approaches by offering rapid and accurate disease diagnosis. Conserve an hour for both doctors and patients as a result. Given that heart illness is currently the biggest reason for demise universally, it ranks among the most difficult illnesses to identify. In this paper, we report an investigation of the machine learning classification methods proposed to help physicians diagnose heart disease. We begin by providing a general review of machine learning and providing detailed descriptions of the most popular classification algorithms for the diagnosis of heart disease. After that, we discuss relevant studies on applying machine learning categorization approaches in this area. A thorough tabular assessment of the examined articles is also provided.

II.INTRODUCTION

Making computers more intelligent is the goal of artificial intelligence (AI), a branch of computer science. Learning is a fundamental condition for intelligence, hence machine learning (ML) emerged as a subfield of AI. Machine learning (ML) is one of the branches of artificial intelligence that is growing the quickest and is used in many areas of daily life, most notably in the healthcare sector. Since ML is a sophisticated tool for data analysis and the medical profession is full of data, it has enormous utility in the healthcare industry. Because of the digital revolution, a lot of data has been gathered and stored in the last few years. Monitoring and other data collecting instruments are widely available, frequently used, and generating enormous amounts of data in today's hospitals. Because it is extremely difficult or even difficult for individuals to extract valuable material from such vast volumes of dataset, machine learning now become frequently employed to examine this data and identify issues in healthcare industry. The machine learning algorithms would acquire from patient instances that had already been diagnosed, to put it simply. The resulting classifier can be used for a variety of objectives, containing teaching students and non-specialists how to diagnose patients and assisting doctors in diagnosing new cases more quickly and effectively.

Machine learning can assist people in identifying arrangements and useful material from the massive volumes of medical datasets that we currently have. In spite of its wide range of uses, machine learning is most commonly applied in the medical field to predict illnesses. Machine learning has attracted the interest of many researchers since it can speed up the diagnosis process while also improving accuracy and efficiency. Machine learning techniques can be used to diagnose a variety of illnesses, however, the analysis of heart disease will be the main topic of this study. Since heart illness is the main reason for demise in the modern earth, an accurate identification of the condition is vital for life preservation [1].

Heart disease, often known as cardiovascular disease, refers to a wide range of conditions that affect the heart. As per the report of the World Health Organization, heart illness is at fault for 12 million deaths worldwide each year. It is the leading killer in many poor nations. For instance, it claims one life every 34 seconds in the US. One of the most serious illnesses threatening adulthood nowadays is heart analysis, as it is also the primary cause of mortality in India [2]. The identification of heart illness is one of the valuable and difficult duties in the sector of the medical industry. To save lives, it needs to be diagnosed effectively, swiftly, and with precision. The patient must undergo a variety of tests, and healthcare specialists must carefully review the results. Because of this, Scientists have developed a number of systems using various machine learning (ML) techniques and have been involved in creating predictions concerning cardiac illness [3]. More of them had successful outcomes than not. While some trained and evaluated their classifiers using data from other nearby hospitals, many others used the well-known UCI heart disease dataset.

An overview of the machine learning classification techniques used in the diagnosis of cardiac diseases will be given in this review paper, along with an account of their previous applications by other researchers. It clarifies the role that machine learning plays in the healthcare sector and shows how it may help doctors by making exact predictions.

This is how the rest of the essay is organized. Basic knowledge on classification techniques, machine learning (ML), and the most popular dataset for heart disease research are covered in Section 2. A summary of the research on the recently recommended search effort in this field may be found in Section 3. The classification algorithms described in Section 3 are tabulated and compared based on correctness in Section 4. Section 5 offers the essay's conclusion.

III. TECHNOLOGIES USED

In this section, the associated topics of this work are described. These subjects cover machine learning, data preprocessing, performance evaluation metrics, and a succinct overview of the most widely used dataset for heart disease.

III.I MACHINE LEARNING

The creation of algorithms with experience-based learning is the focus of the artificial intelligence subfield of machine learning (ML). Aiming to uncover patterns in the input dataset, machine learning algorithms generate models. Next, for newly created datasets that are completely unknown to the algorithms, they are able to produce exact predictions. Learning gave the computer intelligence in this way, allowing it to identify patterns that would be very hard or impossible for people to notice on their own. Machine learning algorithms and methodologies have the capacity to produce assessments and forecasts based on extensive datasets [4]. Figure 1 depicts a basic example of how machine learning works. This picture shows the first preprocessing of the dataset, which in our example may be a patient database. Because preprocessing cleans up the dataset and prepares it for usage by

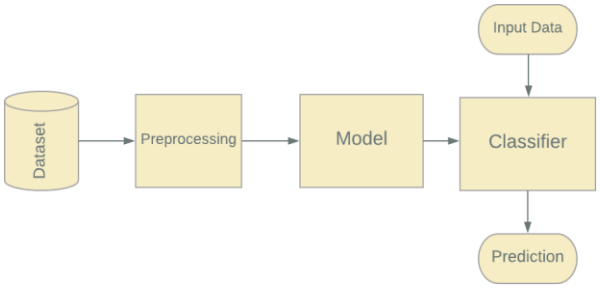
the machine learning algorithm, it is crucial. One algorithm or multiple algorithms working together in a hybrid design could comprise the model. The model's output, or classifier, is where intelligence resides and predictions are generated. The classifier can predict without human intervention if it is given input data. The input data can be new patient information, for instance, if the medical dataset used to feed the model contains data on both healthy and sick patients. The classifier has never encountered any input data like this before. Using this information and previous data, the classifier will determine if the new patient is healthy or not.

III.I Machine Learning Techniques

The following categories describe the essential ML techniques:

III.I.I Supervised Learning

For this method, there is a dataset with examples and their answers (the output). During the training phase, the algorithm can absorb knowledge from the dataset and apply that knowledge to respond to fresh input. Two examples of supervised learning are classification and regression [5].



1.1.1. Unsupervised learning

There are no responses from this method in the dataset. Consequently, the algorithm searches the input data for patterns and clusters the data based on those patterns. One element of the technology for unsupervised learning is the clustering approach [5].

1.1.2. Reinforcement learning

This approach is a hybrid of supervised and unsupervised learning, in which the model learns via its interactions with the outside environment. as a result, learn how to correct its mistakes. After conducting extensive research and trying numerous solutions, the ideal result should be achieved [5].

Especially the classification strategy, which is frequently applied to predicting, supervised learning is the most popular learning methodology. Studies that used classification algorithms to diagnose heart illness are the main focus of this research.

1.2. Machine Learning Techniques for Classification

Classification is a type of supervised machine learning technique that uses past data to predict future events. In this section, the most widely used categorization systems for heart disease prediction are briefly described.

1.2.1. Naïve Bayes, or NB

The Naive Bayes classifier is one of a family of probabilistic classification algorithms based on the Naive Bayes theorem. The assumption of robust feature independence is an essential component of this classifier's prediction process. It can be used for diagnosing illnesses and in the field of medical study because it is simple to assemble and generally functions effectively [6].

1.2.2. Artificial Neural Network (ANN)

In order to replicate the neurons in the human brain, this algorithm was created. It is made up of a number of interconnected nodes or neurons, where one node's output serves as another node's input. Although each node has several inputs, it only has one value to output. A popular kind of ANN is the Multi-Layer Perceptron (MLP), It is composed of an output layer, hidden layers, and an input layer. Every level receives a variable amount of nerve cell depending on the circumstances [6].

1.2.3. RBF, or Radial Basis Function

This ANN kind resembles the Multi-Layer Perceptron (MLP) Neural Network in certain ways, but it differs from it in terms of hidden layer count, approximation method, parameter count as well and additional components [6].

1.2.4. DT, or Decision Tree

This process is structured that follow a tree or a flowchart. Nodes, branches, leaves, and a root node are all present. The properties are kept on internal nodes, and each internal node's test run results are shown on the branches. Since it requires little specific expertise or parameter configuration to operate, DT is frequently used in categorization [6].

1.2.5. KNN, or K-Nearest Neighbor

Using the majority of votes from its close neighbors, this method forecasts the class of the new instance. Euclidean distance is used to compute the distance between an attribute and its neighbors [6].

SVM or Support Vector Machine

The accuracy of this algorithm's classification is useful. It is defined as a finite-dimensional vector space in which every property or characteristic of an object has a distinct dimension [6].

1.2.6. Genetic Algorithm

Based on Darwin's theory of evolution, it is an evolutionary algorithm. It mimics natural processes like natural selection, crossover, and mutation. The use of the evolutionary algorithm to establish the neural network model's weights is among its most popular benefits [8]. Because of this, many studies employ it in conjunction with ANN to create hybrid prediction models.

1.2.7. EL, or Ensemble Learning

Several classifiers are combined into a single model using this strategy to improve accuracy. Three categories of strategies exist for group learning. Bagging is the first type, when classifiers of the same kind are grouped together by a vote process. The other kind is boosting, which is comparable to bagging but takes into account the results of previous models. Stacking [6].

1.3. Data Preprocessing

The quality of the dataset and the preprocessing methods affect the prediction model's performance and accuracy in addition to the algorithms employed. Preprocessing refers to the operations done on the dataset before using the dataset with machine

learning techniques. Because preprocessing gets the dataset ready and converts it into an algorithm-friendly structure, it is essential.
Datasets may have mistakes, omissions, noise, and dismissals, with another issue that makes them inappropriate for use directly by machine learning algorithms. The dataset's magnitude is another consideration. A few datasets contain a lot of characteristics that make it challenging for the algorithm to evaluate them, look for patterns, or anticipate outcomes. By examining the dataset and employing the appropriate data preprocessing procedures, such issues can be resolved. Depending on the structure of the dataset, data preprocessing steps may also include procedures such as feature selection, data normalization, data transformation, data cleaning, and imputation of missing values [9].

1.4. Performance Evaluation Metrics

Researchers assess prediction models and present the results of their performance using the metrics listed below. Without getting into the intricate technicalities and mathematical computations, we give a brief explanation of each strategy.

1. **Accuracy:** This metric displays the proportion of accurated results.
2. **Precision:** This measure shows the degree of relevance of the outcome.
3. **Sensitivity or Recall**: Evaluate the pertinent findings that are returned.
4. **F-Measure:** Combines precision and recall.
5. **ROC or Receiver Operation Characteristic:** This graph shows how well the classifier is doing. Both the correctly classified cases and the wrongly classified cases are displayed. [6].

Every research study examined in our article employs accuracy as its primary performance evaluation parameter. As a result, this overview article's main objective is to classify, contrast, and evaluate prior work based on accuracy.

Heart Disease Dataset

The dataset on heart disease from the Center for Machine Learning and Artificial Intelligence at UCI, located in Irvine, California intelligent systems is the one that is used in the bulk of research papers. It includes four hospital-related databases. There are 14 features in total throughout all databases, however there are various amount of records in each. The Cleveland dataset is the one that machine learning researchers utilize the most because it has more records and has less missing variables than the other datasets. The "num" field indicates whether the patient has cardiac disease. It can have integer values between zero (no presence) to four. There are 303 examples in the Cleveland Information Set [10]. The dataset's 14 attributes/features are listed in Table 1 along with a brief description of each one.

3. Current Classification Techniques for Predicting Heart Disease

Numerous researchers predict heart illness using a range of classification schemes. We give an overview of the surveyed papers in this field in this section. Based on the methods that were employed in each person's prediction model, we categorized each person. The last section, referred to as the "Hybrid Approach" section, contains comparisons of the various algorithms that the majority of researchers combined in their study effort.

Table 1: Dataset Attributes

Number	Attribute	Description
1	Age	Age in years
2	Gender	Male or Female
3	cp	Chest pain type
4	trestbps	Resting blood pressure in mmHg
5	chol	Serum cholesterol in mg/dl
6	fbs	Fasting blood sugar
7	restecg	Resting electrocardiographic results
8	talach	Peak heart rate attained
9	exhang	Angina brought by exercise
10	old peak	Caused ST depression by exercise relative to rest
11	incline	The peak exercise ST segment’s slope
12	ca	Major vessel count (0-3), hued by flourosopy
13	thal	Thallium heart scan
14	num	Diagnosis of heart disease (angiographic disease status)

Naive Bayes

Naive Bayes classifier was used by Vembandasamy et al. in [11] to determine if something is present or not in cardiac illness. One of the top centers for diabetes research in Chennai provided the dataset for the study, which included information for 500 patients with 11 variables (including the diagnosis). The Naive Bayes classifier is applied using the ML techniques are part of the Waikato Environment for Knowledge Analysis (WEKA) program. Their research's accuracy rate was 86.4198%.
In [12], Medhekar et al. introduced a technique that used a Naive Bayes classifier to divide the data into five groups. No, low, average, high, and extremely high are the available categories. In the input data, the algorithm makes predictions about the likelihood of heart disease. The UCI heart disease dataset information, which table 1 illustrates, is the one utilized for training and testing. The accuracy of the system was 88.96%.

3.1. Artificial Neural Network (ANN)

A system utilizing the Ensemble Artificial Neural Network (ANN) approach was proposed by Das et al. [7]. Table 1 displays the Cleveland heart disease dataset. Because it combined several models that had been trained on the same task, the ensemble model boosted generalization. The experiment was carried out using 5.2 SAS Enterprise Miner, as well as the findings revealed the model correctly predicted heart illness 89.01% of the time.

A heart disease prediction system (HDPS) was created by Chen et al. in [13] utilizing an artificial neural network. This study made use of Learning Vector Quanti- zation (LVQ), a kind of ANN. Thirteen neurons were employed in the input layer, six in the secret layer, as well as two in the output layer of the ANN model in this study. The Cleveland dataset in table 1 served as the basis for the study. Users must enter the thirteen medical attributes into the developed system's user-friendly interface in order to generate predictions. The output shows the predicted outcome, ROC curve, accuracy, sensitivity, specificity, and running duration, in addition to whether the condition is healthy or not. The C programming language and C# were used to create the user interface for the system. The findings showed that the model had roughly 80%, 85%, and 70% accuracy, sensitivity, and specificity.
Dangare A Heart Disease Prediction system (HDPS) was created by Dangare and Apte in [14] using ANN to determine whether a patient had heart disease or not. It used the Statlog dataset for testing and the Cleveland heart disease dataset (Table 1) for algorithm training; both were taken consist of thirteen medical attributes and are taken from the UCI library. To boost accuracy, two more attributes—smoking and obesity—were included, bringing the total to fifteen. The WEKA tool is the one utilized for experimentation. The findings indicated that whereas the fifteen traits offered an accuracy of almost 100% for predicting the disease, The accuracy of the thirteen qualities was 99.25%.

Decision Tree (DT)

The Tree of Decisions J48 algorithm was utilized by Sabarinathan and Sugumaran in [15] for feature selection and heart disease prediction. Thirteen medical variables or features were included in the dataset, and 120 records were utilized for testing, while 240 records were used for training. When employing all features, accuracy was 75.83%; however, when using feature selection, accuracy was increased to 76.67%. Furthermore, the accuracy is demonstrated to be 85% when other pointless elements are removed. According to the article, the J48 algorithm enables using the bare minimum of features to improve prediction accuracy.
Using the WEKA tool and the UCI dataset, Patel et al. in [16] tested a number of decision tree methods to identify the presence or absence of cardiac disease. According to the article, Random Forest, the J48 logistic model tree, and many algorithms were put to the test. With an accuracy of 56.76%, the J48 algorithm outperformed the competition.

3.2. K-Nearest Neighbor (KNN)

K-nearest neighbor (KNN) was used by Shouman et al. [17] to predict heart disease using the Cleveland dataset. The results of using KNN alone and KNN in addition to the voting technique were compared in the article. Voting is the procedure of applying the classifier to each group after dividing the data into smaller ones. The evaluation method is tenfold cross-validation. Based on the value of K, the accuracy ranged from 94% to 97.4% in the absence of voting, according to the results. When K=7, the accuracy was at its highest, 97.4%. However, using the voting method did not increase the accuracy. The findings revealed that the accuracy fell to 92.7% at K is eqal to 7.

3.3. SVM, or Support Vector Machine

In [18], Using the UCI dataset, Wiharto et al. examined the diagnostic performance of several SVM algorithm types. Decision Direct Acyclic Graph (DDAG), One-Against-One (OAO), One-Against-All (OAA), and Exhaustive Output Error Correction Code (ECOC) are only a few of the SVM types used in the study. Initially, the dataset was pre-processed using a min-max scaler. Using the previously mentioned SVM techniques, the algorithm was then trained on the dataset. With an overall accuracy of 61.86%, BTSVM performed better in the performance evaluation than the other algorithms.

3.4. Approach of Hybrid

In this segment includes studies that developed a model that employs several algorithms or that compared various algorithms.
On the UCI Cleveland dataset, Khateeb and Usman in [3] tested a number of classification techniques, including Naive Bayes, KNN, decision trees, and bagging approach. The job was separated into 6 cases, and each classifier calculated The precision concerning each case individually. In example 1, there was no feature reduction and every classifier was used on the given dataset. In example 2, feature decrease was used, and just seven of the dataset's 14 attributes—the ones that are most crucial for diagnosing heart disease—were chosen. Only the most universal characteristics, such as blood sugar levels during rest, age, and sex, were eliminated in Case 3. In case 4, only the seven most important attributes were used, and the WEKA tool resampled the dataset to improve the precision of every classifier. In case 5, all 14 attributes were resampled, and in case 6, The WEKA tool made use of the Synthetic Minority Over-sampling Technique (SMOTE). With case 5, KNN yielded the greatest result, producing an accuracy of 79.20%.
A comprehensive examination of different categorization techniques was conducted by Pouriyeh et al. in [6] using the Cleveland heart disease dataset to find the classifier that performs better than the others. Single Conjunctive Rule Learner (SCRL), Radial Basis Function (RBF), K-Nearest Neighbor (KNN), Decision Tree (DT), Naive Bayes (NB), Multi-layer Perceptron (MLP), and Support Vector Machine (SVM) were some of the classifiers that were included. The comparison of ensemble strategies including bagging, boosting, and stacking was also covered in the paper. The authors computed the classifiers' accuracy using the K-Fold Cross Validation method, which yields accuracy, precision, and recall, the F-measure, & the ROC curve were the performance evaluation metrics for each classifier. Different values of K were evaluated for the KNN classifier, and K=9 was found to be the optimum value. The optimal combination for an ANN, which is the input, hidden, and output layers' respective numbers: thirteen, seven, and two was determined by testing different neuron numbers. The study was split into two experiments: the first compared the various classifiers listed above, and the second used ensemble approaches. SVM performed better than84.15 percent accuracy compared to the other classifiers in the first experiment, based on data. Employing the boosting technique with SVM also proved to be the most effective in the second experiment, with an accuracy of 84.81%.
Heart illness using a hybrid system prediction utilizing a Genetic algorithm ANN was proposed by Amin et al. [19]. The American Heart Association performed a survey of 50 persons to create the dataset used in this study, and it included thirteen features. Preprocessing the data before analysis involved

eradicating any invalid or missing values. The dataset was split into two parts: 15% for testing and validation and 70% for training. MATLAB R2012a was used to develop the system using the Neural Network Toolbox and Global Optimization Toolbox. The results showed that an 89% accuracy rate may be used to diagnose cardiac disease in an individual.

A cardiac disease prediction system was created by Waghulde and Patil [8] utilizing a genetic algorithm and ANN. The weights in the neural network were initialized using a genetic approach. Thirteen attributes were included in the trial, which was conducted by applying MATLAB on a 50-piece dataset of individuals gathered by the American Health Association. When six hidden nodes and 10 hidden nodes were used, the accuracy of the results was 98% and 84%, respectively.

Amma [20] described a system that combines an ANN with a genetic algorithm to diagnose heart disease. The Cleveland dataset was the one that was used. The dataset was preprocessed by adding missing values and applying Min-Max normalization to the data. The genetic method was used to determine the neural network's weights. The obtained accuracy was 94.17%. Naive Bayes and Decision Tree were compared by Venkatalakshmi and Shivsankar [21] to see which one predicts cardiac illness with the most degree of accuracy. The cardiac illness dataset from UCI was the one that was used. The experiment, which used the WEKA tool, revealed accuracy for the Naive Bayes and Decision Tree models of 85.03% and 84.01%, respectively. Prior to using the WEKA tool to process the dataset, further research, the report recommended employing using a genetic algorithm in MATLAB to reduce the total number of features.

A Neural Network, Decision Tree, and Naive Bayes-based Intelligent Heart Disease Prediction System (IHDPS) was proposed by Palaniappan and Awang [22]. The.NET framework was used to construct the web-based system. The Cleveland Heart Disease database's 909 records with 15 attributes made up the data source. The model was developed using the Data Mining Extension (DMX) query language. The findings indicated that Naive Bayes was the most effective model,Neural Network trailed by only 1%, with 86.53% of predictions made correctly.

A cardiac disease prediction model was created by Dangare and Apte [23]. The dataset consists of 303 records from the Cleveland database and 270 records from the Statlog database. In addition to the thirteen already present features in the dataset, they added two more: smoking and obesity. The dataset was preprocessed using the WEKA tool. The dataset was analyzed using three classification methods: ANN, Decision Tree, and Naive Bayes. ANN accuracy was 100%, Decision Tree accuracy was 99.62%, and Naive Bayes accuracy was 90.74%, demonstrating that Artificial Neural Network is the best algorithm.

An efficient intelligent medical decision support system was created by Zriqat et al. [24]. The classification techniques of Random Forest, Discriminant, Naive Bayes, Decision Tree, and Support Vector Machine were contrasted. The Statlog Heart Disease and the Cleveland Heart Disease datasets are two and were subjected to analysis using MATLAB. Using the Cleveland and Statelog datasets, respectively, the decision tree achieved the highest accuracy for both datasets, according to the results, at 99.01% and 98.15%.

In [25], Liu et al. suggested a hybrid methodology for identifying cardiac illness. UCI's Statlog heart illness dataset library was the one utilized. The two subsystems of the MATLAB-created model were the selection and categorization of features. The component for feature selection estimates the weight of features using the Relief technique before removing extraneous features and enhancing model accuracy using the Rough Set method (RFRS). Classification subsystem was based on Ensemble classifier with C4.5 approach (which generates a Decision Tree). The categorization accuracy rate was 92.59%, according to the results.

RBF or Radial Basis Function, a kind of ANN, & the Support Vector Machine were contrasted by Ghumbre et al. [26]. The algorithms were used to determine whether or not a person had heart disease on 214 records and 19 characteristics in an Indian patient dataset. Using the dataset for training and testing, two cross-validation methods: five-fold and ten-fold were used to assess the algorithms' performance. For SVM and RBF, accuracy of 86.42% and 80.81% were generated by the total average performance, respectively. Their findings demonstrated that SVM demonstrated superior accuracy.

To identify cardiac illness, Masethe and Masethe in [27] used a number of algorithms, including Simple Cart (Classification and Regression Tree), Naive Bayes, J48, a Decision Tree variant, and Bayes Net. The study's dataset, which came from South African physicians, had the eleven characteristics listed below: Patient information (replaced with dummy values to preserve patients' privacy), gender, age, blood pressure, heart rate, cholesterol, smoking, and alcohol intake, as well as symptoms of chest pain. The WEKA tool was the instrument utilized in the experiment. In order to evaluate the effectiveness of the developed model. We used 10-fold cross-validation to evaluate performance. A comparison of the accuracy values for J48, REPTREE, 97.222%, 98.1481%, and the basic cart shows that it fared better than the others. The simple cart's accuracy was 99.0471%.

4. Evaluation of ML Classification Methods for Predicting Heart Disease

In this part, all of the research papers mentioned above are compared in a table.

Table 2 shows that accuracy is the basis for comparison. The following six elements make up the table::

1. **Author:** This displays the paper's authors and their reference number.
2. **Classification Technique/s:** Regardless of whether it was a comparison, a single algorithm, or a hybrid model, this reflects the categorization algorithm that was employed in the study.
3. **Best Technique Found**: Papers that compare various algorithms are the only ones that can use this column. That stands for the most accurate algorithm discovered during the research project.
4. **Tool:** This column displays the framework or programming language that was used to create the model. The prediction model was developed, **and** tested, and the input dataset was pre-processed using it.
5. **Dataset:** This displays the dataset that served as the classification algorithm's input.
6. **Accuracy:** This shows how accurate the outcomes of the suggested model were. This column only shows the best approach used by the author in terms of accuracy, if there was a comparison in the

publication.

5. Wrapping Up and Closing Thoughts

In this study, the literature on machine learning classification algorithms for diagnosing heart disease is examined. We reviewed and categorized a large number of representational publications using machine-learning classification approaches. The classifier employed in the model, the preprocessing methods, the instrument used, the dataset used, and the number of attributes and records in the dataset, all affect how accurate the proposed models are. Whether or not the model employs feature selection and whether or not it is a hybrid model will determine this. Table 2 shows that Dangare and Apte achieved the most accurate results utilizing the WEKA tool, an Artificial Neural Network (ANN), and a combination of the Cleveland and Statlog heart disease datasets.

To develop an accurate heart disease prediction model, we conclude that a dataset with sufficient samples and trustworthy data must be used. The preprocessing of the dataset is the most important step in getting the best results from the machine learning algorithm, hence it must be done correctly. The creation of a prediction model also requires the usage of an appropriate algorithm. Based on the majority of models for heart illness prediction, it is evident that both Decision Trees (DT) and Artificial Neural Networks (ANN) performed well.

Lastly, using machine learning to diagnose cardiac disease is an important topic that benefits people as well as medical professionals. Because the field is still evolving, not much of the massive amount of patient data that is available in hospitals and clinics gets published. As can be seen in Table 2, the UCI repository is where the majority of researchers obtained their datasets. Since the dataset's quality plays a crucial role in how accurately a prediction is made, encouragement should be given to additional hospitals to supply high-quality datasets (while protecting patient privacy), giving researchers a reliable resource from which to build their models and get good results.

REFRENCES:

[1] P. K. Kushwaha and M. Kumaresan, "Machine learning algorithm in healthcare system: A Review," 2021 International Conference on Technological Advancements and Innovations (ICTAI), Tashkent, Uzbekistan, 2021, pp. 478-481, doi: 10.1109/ICTAI53825.2021.9673220.

[2] P. K. Kushwaha, B. P. Lohani and D. Singh, "Review on information security, laws and ethical issues with online financial system," 2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH), Greater Noida, India, 2016, pp. 49-53, doi: 10.1109/ICICCS.2016.7542350.

[3] G. Gulati, B. P. Lohani and P. K. Kushwaha, "A Novel Application Of IoT In Empowering Women Safety Using GPS Tracking Module," 2020 Research, Innovation, Knowledge Management and Technology Application for Business Sustainability (INBUSH), Greater Noida, India, 2020, pp. 131-137, doi: 10.1109/INBUSH46973.2020.9392193.

[4] D. Pareta, I. N. Verma, B. P. Lohani, P. K. Kushwaha and V. Bibhu, "IoT Enabled Smart and Efficient Musical Water Fountain," 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), Gautam Buddha Nagar, India, 2022, pp. 369-373, doi: 10.1109/ICIPTM54933.2022.9754129.

[5] B. P. Lohani, M. Trivedi, R. J. Singh, V. Bibhu, S. Ranjan and P. K. Kushwaha, "Machine Learning Based Model for Prediction of Loan Approval," 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2022, pp. 465-470, doi: 10.1109/ICIEM54221.2022.9853160.

[6] A. Kumar, B. P. Lohani and P. K. Kushwaha, "Robust Secured Framework for Online Business Transactions over Public Network," 2021 2nd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2021, pp. 555-560, doi: 10.1109/ICIEM51511.2021.9445380.

[7] P. K. Kushwaha and B. P. Lohani, "A review of security of the cloud computing over business with implementation," 2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH), Greater Noida, India, 2016, pp. 192-198, doi: 10.1109/ICICCS.2016.7542342.

[8] M. Chandra, P. K. Kushwaha and S. Saxena, "Modified Fractal Carpets," 2011 International Conference on Computational Intelligence and Communication Networks, Gwalior, India, 2011, pp. 537-540, doi: 10.1109/CICN.2011.115.

[9] P. K. Kushwaha, R. Kohli and D. Singh, "Secret key watermarking in WAV audio file in perceptual domain," 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), Greater Noida, India, 2015, pp. 629-634, doi: 10.1109/ABLAZE.2015.7154940.

[10] Ranjan, Ankur A. et al. "An Approach for Netflix Recommendation System using Singular Value Decomposition." Journal of Computer and Mathematical Sciences (2019)

[11] Makkar, Bhavya et al. "Map Reduce concept-based Sentiment Analysis Approach." International Journal of Computer Sciences and Engineering (2019)

[12] Bhatia, Ayush & Bibhu, Vimal & Lohani, Bhanu & Kushwaha, Pradeep. (2020). An Application Framework for Quantum Computing using Artificial intelligence Techniques. 264-269. 10.1109/INBUSH46973.2020.9392164.

[13] A. Kumar, B. P. Lohani and P. K. Kushwaha, "Black Hole Attack in Mobile Ad Hoc Network and its Avoidance," 2021 International Conference on Innovative Practices in Technology and Management (ICIPTM), Noida, India, 2021, pp. 103-107, doi: 10.1109/ICIPTM52218.2021.9388366.

[14] Srivastav, A.V., Lohani, B.P., Kushwaha, P.K., Tyagi, S. (2021). Dual-Layer Security and Access System to Prevent the Spread of COVID-19. In: Prateek, M., Singh, T.P., Choudhury, T., Pandey, H.M., Gia Nhu, N. (eds) Proceedings of International Conference on Machine Intelligence and Data Science Applications. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-33-4087-9_28

[15] A. Khuran, B. P. Lohani, V. Bibhu and P. K. Kushwaha, "An AI Integrated Face Detection System for Biometric Attendance Management," 2021 2nd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2021, pp. 29-33, doi: 10.1109/ICIEM51511.2021.9445295.

[16] S. Salagrama, B. P. Lohani and P. K. Kushwaha, "An Analytical Survey of User Privacy on Social Media Platform," 2021 International Conference on Technological Advancements and Innovations (ICTAI), Tashkent, Uzbekistan, 2021, pp. 173-176, doi: 10.1109/ICTAI53825.2021.9673402.

[17] S. Singh, D. Chaudhary, A. D. Gupta, B. Prakash Lohani, P. K. Kushwaha and V. Bibhu, "Artificial Intelligence, Cognitive Robotics and Nature of Consciousness," 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2022, pp. 447-454, doi: 10.1109/ICIEM54221.2022.9853081.

[18] S. Suman, P. Kaushik, S. S. N. Challapalli, B. P. Lohani, P. Kushwaha and A. D. Gupta, "Commodity Price Prediction for making informed Decisions while trading using Long Short-Term Memory (LSTM) Algorithm," 2022 5th International Conference on Contemporary Computing and Informatics

(IC3I), Uttar Pradesh, India, 2022, pp. 406-411, doi: 10.1109/IC3I56241.2022.10072626.
[19] P. William, Y. V. U. Kiran, A. Rana, D. Gangodkar, I. Khan and K. Ashutosh, "Design and Implementation of IoT based Framework for Air Quality Sensing and Monitoring," 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2022, pp. 197-200, doi: 10.1109/ICTACS56270.2022.9988646.
[20] Mridul Bhardwaj and Ajay Rana. 2015. Impact of Size and Productivity on Testing and Rework Efforts for Web-based Development Projects. SIGSOFT Softw. Eng. Notes 40, 2 (March 2015), 1–4. https://doi.org/10.1145/2735399.2735404
[21] Bhardwaj, Mridul, and Rana Ajay. "Estimation of testing and rework efforts for software development projects." Asian Journal of Computer Science and Information Technology 5.5 (2015): 33-37.
[22] Dubey, Gaurav, Ajay Rana, and Jayanthi Ranjan. "A research study of sentiment analysis and various techniques of sentiment classification." International Journal of Data Analysis Techniques and Strategies 8.2 (2016): 122-142.
[23] R. Sharma, M. Mogha, S. Tanwar and A. Rana, "Emerging Part of Industry 4.0: The Digital and Physical Technology," 2020 9th International Conference System Modeling and Advancement in Research Trends (SMART), Moradabad, India, 2020, pp. 149-154, doi: 10.1109/SMART50582.2020.9337064.
[24] Dubey, Sanjay Kumar, and Ajay Rana. "Assessment of usability metrics for object-oriented software system." ACM SIGSOFT Software Engineering Notes 35.6 (2010): 1-4.
[25] Singh, Archana, Jyoti Agarwal, and Ajay Rana. "Performance Measure of Similis and FPGrowth Algo rithm." International Journal of Computer Applications 62.6 (2013): 25-31.
[26] Tyagi, Neha, Ajay Rana, and Vineet Kansal. "Load distribution challenges with virtual computing." Intelligent Computing in Engineering: Select Proceedings of RICE 2019. Springer Singapore, 2020.
[27] Singh, Jaya, and Ajay Rana. "Exploring the big data spectrum." International Journal of Emerging Technology and Advanced Engineering 73 (2013).
[28] N. M., P. Chawla and A. Rana, "A Practitioner's Approach to Assess the WCAG 2.0 Website Accessibility Challenges," 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 2019, pp. 958-966, doi: 10.1109/AICAI.2019.8701320.
[29] Tyagi, N., Rana, A., Awasthi, S., & Tyagi, L. K. (2022). Data Science: Concern for Credit Card Scam with Artificial Intelligence. In Cyber Security in Intelligent Computing and Communications (pp. 115-128). Singapore: Springer Singapore.
[30] Jain, Piyush, Sanjay Kumar Dubey, and Ajay Rana. "Software usability evaluation method." International Journal of Advanced Research in Computer Engineering & Technology 1.2 (2012): 28-33.
[31] Pal, S. K., et al. "Hanging suicides in himachal pradesh: an analysis of forensic cases." Int J Forensic Sci Pathol 4.11 (2016): 297-304.
[32] Rana, A., and S. Manhas. "Significance of diatoms in diagnosis of drowning deaths: a review." Journal of Forensic & Genetic Sciences 5 (2018): 77-81.
[33] Bansal, Sangeeta, and Dr Ajay Rana. "Transitioning from relational databases to big data." International Journal of Advanced Research in Computer Science and Software Engineering 4.1 (2014): 394-400.