

Application of Machine Learning for the Detection of Heart Disease

Samir S Yadav

Department of Information Technology
Dr. Babasaheb Ambedkar Technological University, Lonere
Raigad, India
ssyadav@dbatu.ac.in

Shivajirao M. Jadhav

Department of Information Technology
Dr. Babasaheb Ambedkar Technological University, Lonere
Raigad, India
smjadhav@dbatu.ac.in

Snigdha Nagrale

Department of Information Technology
Dr. Babasaheb Ambedkar Technological University,
Lonere
Raigad, India
13snigdhanagrale@gmail.com

Niraj Patil

Department of Information Technology
Dr. Babasaheb Ambedkar Technological University,
Lonere
Raigad, India
patilniraj67@gmail.com

Abstract—One of the most common tasks in machine learning is data classification. Machine learning emerges as a key feature to derive information from corporate operating datasets to large databases. Machine Learning in medical health care is evolving as a significant research field for delivering prognosis and a deeper understanding on medical data. Most methods of machine learning depend on several features defining the behavior of the algorithm, influencing the output, and the complexity of the resulting models either directly or indirectly. Many machine learning methods have been used in the past to detect heart diseases. Neural network and logistic regression are some of the few popular machine learning methods used in heart disease diagnosis. They analyze multiple algorithms such as neural network, K-nearest neighbor, naive bayes, and logistic regression along with composite approaches incorporating the aforementioned heart disease diagnostic algorithms. The system was implemented and trained in the python platform by using the UCI machine learning repository benchmark dataset. For the new data collection, the framework can be extended.

Index Terms—UCI machine learning repository, K-nearest neighbors, neural network, logistic regression

I. INTRODUCTION

Heart disease (HD) or cardiovascular disease is the most common human disease in the world that imposes a threat to life. In this state, the heart is usually unable to pump the required amount of blood to various parts of the human body during order to meet the regular functionalities of the body, resulting in irreversible heart failure. The United States has an extremely high rate of heart disease [1]. Symptoms such as breathlessness, physical body weakness swollen ankles and tiredness with the associated sign, such as increased vein stress and peripheral edema from functional or non-cardiac

disorders, are symptomatic of cardiac illnesses. Early-stage testing strategies for the diagnosis of HD have been difficult and their complexity is one of the main reasons for impacting the quality of living [2]. Diagnosis and treatment of cardiovascular diseases are extremely complex, especially in developing countries. Since diagnostic instruments are rarely available and medical practitioners and other missing services may affect the correct prediction and care of cardiac patients [3].

A lot of research and study has been done over the last few years on deploying better and reliable data sets for heart disease. [4] offers a strategy based on knowledge. Initially, Dr. Robert Detrano used logistic regression to achieve a precision of 77% [5]. Newton Cheung used algorithms with Naive Bayes and obtained a classification accuracy of 81.48% [6]. In treating patients with heart disease, Palaniappan and Awang explored the analysis of various data mining techniques. Naive Bayes, decision trees, and neural networks were used in these techniques. The results showed that Naive Bayes was able to achieve the highest precision in diagnosing patients with heart disease [7].

Indira [8] uses the BF network of probabilistic artificial neural networks. This is a radial-based algorithm class RB approach is useful for automated pattern recognition, nonlinear modeling, and class composition probability estimation and probability ratios. With a total of 576 records and 13 patient features, the data used in the tests were drawn from the Cleveland heart disease archive. The best performance in accuracy was 94.60%. Among the most widely used methods for data mining in grouping issues is KNearest Neighbor [9]. It is a popular

choice due to its simplicity and relatively high level of convergence. Nonetheless, the huge memory space required to store the entire sample is a major disadvantage of KNN classifiers. If the sample is large there is also a large response time on a sequential computer [10]. For regional developmental methods and good statistical precision, the Cleveland dataset used. The research utilized techniques for choosing characteristics. The classification efficiency of the method thus relies on the characteristics chosen. Gudadhe et al. [11] employed multilayer perceptron(MLP) and heart disease classification using support vector machine (SVM) algorithms. They recommended a method of classification and reached 80.41% precision. Python and machine learning have been used in our research to support libraries such as sci-kit learn, numpy, pandas and matplotlib. Some physicians utilized 10-fold cross-validation of the overall data to diagnose medical data, reporting the findings for detection of the condition, while others would not use this method to prevent heart disease. For the optimum set of parameters regarding our analysis, we used both the test train division principle together with cross-validation.

II. MATERIALS AND METHODS

The goals of the research are to perform a” comprehensive analysis of different algorithms in machine learning and develop a more accurate algorithm.” People these days’ work for hours and hours on machines, they have no time to take care of themselves. The poor health of people is severely affected because of the unhealthy lifestyle, hexagonal habits, and intake of junk food. It can determine how much a person would die from heart disease with the medical parameters, which are adequate patients as well as normal people.

The task is to determine which studies will be positive and which will be inaccurate in the cardiovascular identification process. The following subsections address extensively the paper’s analysis materials and procedures.

A. Data set Description

Various studies use” Cleveland heart disease data set 2016” [12] and can be downloaded from the University of California, Irvine’s online data mining depot. This data set has been used in this study to develop a machine learning algorithm based system for the diagnosis of heart disease. The data set for Cleveland cardiac disease is 303 cases with 76 attributes and certain incomplete attributes. During the study, six samples were rejected

due to a missed value of the task columns and a resulting sample size of 297 were taken with 13 further separate input features. There are two groups on the goal performance label to describe a heart patient or a regular topic. The dataset derived is, therefore, a vector of 297 x13. The full description of 297 instances of 13 data sets is shown in Table I.

B. Classification Problem

Throughout Machine learning the question is to classify, on the basis of a training data set of observers whose group membership is identified, which of a number of categories belong to a new observation. Classification is a supervised learning problem, i.e. problems with poorly labeled tuples and algorithms. A classification algorithm is classified as a classifier in particular in an actual implementation [13].

C. Binary and Multiclass Classification

Classification can be viewed as two distinct binary and multi-class classification problems. A better understood function of binary classification needs only two types, while multi type class classification includes the assigning of an entity to one of various classes. Since a large number of classification

Sr. NO.	Attribute	Description	Values
1	Age	Age in Years	Continuous
2	Gender	Male or Female	0= male, 1= female
3	Cp	Chest Pain Variety	4 = typical type 3 = typical type angina 2 = nonangina pain 1 = asymptotic
4	Threst bps	Sleeping Blood Force	Continuous
5	Chloe	Serum Cholesterol	Continuous 5 = normal 4 = having ST T wave abnormal
6	Rest Ecg	Resting ECG	3 = left ventricular hypertrophy
7	Fbs	Fasting Blood Sugar	1 >= 120 mg/dl 0 <= 120 mg/dl
8	Thalach	Greatest heart speed reached	Continuous
9	Ex ang	Exercise induced angina	1 = no 2 = yes
10	Old peak	ST despair induced by apply virtual to relax	Continuous
11	Slope	Slope of the height effect ST section	3 = un sloping 2 = flat 1 = downsloping
12	Ca	Number of key vessels painted by fluoroscopy	1-4 value
13	Thal	Desert Category	2 = normal 4 = fixed 5 = reversible defect

TABLE I HEART DISEASE DATASET(CLEVELAND AND STATLOG)

procedures were originally designed for binary classification; the simultaneous use of multiple binary classifications also requires multiple classifications.

D. Comparison of methods of classification

The methods of classification can be analyzed and compared with the following criteria:

- 1) **Accuracy:** A labeling device's accuracy relates to its ability to properly interpret a tuple or unlabeled data object. The stronger the accurate the algorithm will be.
- 2) **Speed:** Classifier speed refers to the amount of classification time required by the classification algorithm.
- 3) **Robust:** a classifier's sturdiness is its capacity to distinguish accurately when the data are chaotic and outliers.
- 4) **Scalability:** Scalability of a classifier is its ability to classify efficiently when large amounts of data is given.
- 5) **Interpretability:** It applies to the degree that the classifier offers for comprehension and perspective.

E. Machine Learning Classifiers

Machine learning recognition systems are used to identify heart patients and healthy people. In the present paper, we address briefly some popular classification algorithms and their theoretical context.

- 1) **Logistic Regression:** In terms of its name, logistic regression is not a regression model but a linear classification. Logistic regression is also recognized in literature as logistical regression, maximum-entropy classification (MaxEnt). In this model, a logistic function is used to predict the probabilities that characterize the future consequence of a single test. The logistic regression in scientific learning can be approached from the Logistic Regression framework. For optional L2 or L1 regularization this design will work in a multi-class (one vs rest) logistic regression. In order to optimize the logistic regression penalized to binary class L2, the following risk function minimizes:

$$\min_{\omega, c} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^n \log(\exp(-y_i(X_i \omega + c)) + 1)$$

Likewise, the L1 regularized logistic regression addresses the following problem of optimization:

$$\min_{\omega, c} \|\omega\|_1 + C \sum_{i=1}^n \log(\exp(-y_i(X_i \omega + c)) + 1)$$

A logistical regression description may begin with a logistics feature explanation. The logistical function is valuable because its value is always between zero and one and can therefore only be viewed as a probability. It can take any amount from negative to positive infinities. This describes the logistic function as follows:

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{e^{-t} + 1}$$

- 2) **K-means Clustering:** In order to separate samples in n classes that have the same variance, the K-Means algorithm cluster results, minimizing the inertia criterion or using a square number. The number of clusters to be defined is needed for this algorithm. This applies well to a huge number of samples and is used in several different areas of use. The k-means algorithm partitions a set of observations N through K disjoint clusters C. Each is the average sample of μ_j in the cluster. The mean square often referred to as "centroids," usually they are not X points, although they reside in the same space. The K-means algorithm attempts to identify centers within a cluster sum that reduces inertia or squared criteria.

$$\sum_{i=1}^n \min_{\mu_j \in C} (\|x_j - \mu_i\|^2)$$

Inertia can be recognized as a calculation of how internal stability clusters is in the cluster total of the criteria for squares. There are several disadvantages:

- Inertia assumes clusters which is not always the case, are convex and isotropic. It does not react adequately to extended clusters or multiples of uneven structures.
- Inertia is not a measure that has been normalized: we realize that lower and negative values are better. Nevertheless, Euclidean distances appear to become overflowing in very high dimensional spaces (the "curse of depth" is an example). If a dimensional algorithm is used before K-means clustering, this issue could be solved and calculations accelerated.

K-means are often called the algorithm of Lloyd. The method has three stages in basic terms. The first stage is to choose the initial centering, with K samples from the K dataset as the most basic method. K-means is the sequence between the two other phases following initialization. Each sample is allocated to its closest location. In the second stage new centers, all measurements allocated to

each previous center, are calculated in average. The difference is measured between the old and the current centers; and the process takes up to less than one of the last two measurements. With ample time, K-means would still converge, but at least local time. The initialization of centroids is highly dependent. As a consequence, measurements are often carried out multiple times, with various centroid initializations.

- 3) **Naive Bayes:** It is a supervised algorithms collection focused on the theories of the implementation of Bayes with the "naive" presumption that each pair of features is independent. The Bayes theorem states that a class variable y and a dependent x_1 through a x_n feature have been defined:

$$P(y|x_1, \dots, x_n) = \frac{p(y)P(x_1, \dots, x_n|y)}{P(x_1, \dots, x_n)}$$

With the naive assumption of independence

$$P(y|x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = P(x_i|y)$$

This relationship has been simplified for all

$$P(y|x_1, \dots, x_n) = \frac{p(y) \prod_{i=1}^n P(x_i|y)}{P(x_1, \dots, x_n)}$$

As the input is constantly given by $P(x_1, \dots, x_n)$, the following classifying rule can be used:

$$P(y|x_1, \dots, x_n) \propto p(y) \prod_{i=1}^n P(x_i|y)$$

$$y^* = \underset{=1}{\operatorname{argmax}}^y \propto p(y) \prod_{i=1}^n P(x_i|y)$$

and we will use the approximation for $P(y)$ so $P(x_i|y)$ for Maximum a Posteriori (MAP) to measure the relative class Y frequency in the training sets, then the former. The various Naive Bayes classification schemes vary mostly because they believe that $P(x_i|y)$ will be spread. Though it seems to be oversimplified, simplistic Bayes classification schemes have performed very well in many real-world situations, such as text classification and spam filtering. They need a small amount of training data to determine the parameters required. In comparison to more advanced methods, Naive Bayes learners and classifiers can be extremely quick. The disconnection of the class conditional distribution enables each distribution to be calculated separately as a one-dimensional distribution. In addition, this helps reduce problems arising from the dimensionality curse.

- 4) **K-Nearest Neighbors (kNN):** kNN is used for classification and inference by a non-parametric approach of pattern recognition. The input is the closest example

of k in the feature space in both cases. kNN implies an instance-based learning or lazy learning that only approximates the function locally and delays all calculations until classification is completed. The KNN algorithm is one of the easiest algorithms to learn. Also for grouping as well as for estimation, weight can be assigned to the neighbors' inputs so that the near neighbors create a more reasonable contribution than the farther ones. A typical weighting scheme is to assign a weight of $1/d$ for each neighbor, where d is the distance from the neighbor. The underlying objects are drawn from a collection of items defined for the class (kNN classification) or the property value for the object (kNN regression). This can be seen as the algorithm training set but no specific training is required. The kNN algorithm has a weakness because it is sensitive to the local data structure. Another common machine learning method, the algorithm must not be mistaken with K-means.

- 5) **Neural Network:** Machine Learning incorporates the most widely utilized form of the neural network, feedforwards artificial neural networks or, more simply, multilayer perceptron (MLPs). The MLP contains an input layer, output layer and one or several hidden layers. The growing MLP layer comprises of one or more neurons synchronized with neuron layers prior and next. Through MLP, each neuron is the same. Each has multiple input connections and multiple output connections. The values from previous levels, separately for each neuron, and the biological term are summed up in certain weights. The total is converted by the activation function f , which for multiple neurons may be specific. That is, provided x_j in n line, y_i in $n + 1$ layer outputs are determined as:

$$u_i = X(w_{i,jn+1} * x_j) + (w_{i,biasn+1} y_i) = f(u_i)$$

Various functions can be used for activation. Machine Learning has three standard features:

- Identity function : $f(x) = x$
- Symmetrical sigmoid : $f(x) = \beta * (1 - e^{-\alpha x}) / (1 - e^{-\alpha})$, is the default MLP choice.
- Gaussian function : $f(x) = \beta e^{-\alpha x^2}$, which is currently not completely supported.

The activation functions of both neurons in the machine learning process have the same free

parameter with (α, β) specified by a user and not determined by the training algorithms. To measure the network, all w_{ij}^{n+1} weights are required. The training algorithm computes the weights. The algorithm receives a training set, includes multiple input vectors and iteratively adjusts the weights so that the network can respond to the supplied input vectors. The larger the network size, the more robust the system becomes. Arbitrarily, the loss in the training set could be high. But the network learns the noise in the training set as well, so it usually begins to grow after the network size is limited. The errors in the test set are usually increased. Furthermore, larger networks are much longer trained than smaller networks, so that only essential features can be prepared and a smaller network formed. Another MLP feature is that categorical data can -not be handled in the manner it is. But a solution is available. When a function on the input or output (for n classification for $n > 2$) a categorical layer will take $M > 2$ different values, a binary tuple of the M objects can be interpreted where this is 1 and only if the feature is equivalent to the i th value from M . It increases the input/output layer size, but speeds up the convergence of the workout algorithm and at the same time enables "substantial" variables of this type, which means tuples of probabilities rather than fixed values. Two MLP training algorithms are used for machine learning. A standard random algorithm of back propagation is the first algorithm. The latter (default) is an RPROP batch algorithm.

Overfitting and Regularization: Over-fitting is a major problem in neural networks. This is especially true in modern networks, which often have very large numbers of weights and biases. We need a way to detect how to train effectively when over-fitting is going on, so we don't over-train. to over-train. And we want strategies to reduce the overfitting results. Regularization is one way of combating overfitting. The regulation changes the target function by adding additional conditions to penalize large weights, which we minimize. In other terms, we are shifting the target to the Error+ $\lambda \text{d}f(\theta)$, where the power of regularization (a learning algorithm hyper-parameter), grows larger as the components of θ grown larger and λ becomes larger. A $\lambda = 0$ means that the possibility of overfitting is not taken. If λ is too high, our model will focus on keeping θ the parameter value as small as possible when trying to find the values which are good for our workout.

Consequently, λ is a very important task and some trial and error may be required.

- 6) **Fuzzy K- Nearest Neighbors** The method is based on assigning membership to the available groups depending upon the difference between the function and its closest neighbors and those neighbors. In that it still requires to scan the sample set labeled for the K-nearest neighbors, the fuzzy algorithm is identical to the crisp variant. The methods vary considerably beyond the collection of the K samples. While the fuzzy K-nearest neighbor technique often reflects a rating algorithm, the results are different from the flat form. A class affiliation refers to a display function instead of a particular class the Fuzzy K-nearest neighbor algorithm. The drawbacks are that the program does not execute random tasks. Furthermore, the membership values of the variable should provide a confirmation standard that facilitates the subsequent classification. Large memory space is required to store the all the training data. It is also computationally expensive and has a greater time complexity.
- 7) **K Means Clustering with Naive Bayes Classifier** We initially implemented K Means individually. Then we worked on hybrid by combining K Means with naive bayes. K Means is used to group together similar data. Then nave bayes was implemented on each cluster and model was made. For each new test case, it was first determined to which cluster it belongs. Then the naive bayes model for that particular cluster was used to make prediction for the given test case. K Means was used with the hope that grouping together similar data will help in increasing accuracy of naive bayes algorithm. Here we had a tradeoff between time of computation and accuracy but additional gained accuracy was preferred. For this algorithm, we first discretized the data as naive bayes requires data which is in discrete form. We could not use gaussian Naive Bayes because the distribution of data was not gaussian and it would have contributed to poor accuracy if anything were used. We had two choices to discretize records, equal discretizing width and equal discretizing frequency. Equal variable widths contributed to improved performance.

F. CROSS VALIDATION

Cross-validation, also referred to as the estimation of rotation, is a testing technique to evaluate whether mathematical observational findings are translated to an individual set of data. It is primarily used in quantitative environments and one needs to approximate the actual

output of a predictive model. A model generally contains a dataset of known data (training data set), a dataset of unknown data (or first used data) against which the model is evaluated (testing dataset). In a prediction problem, a model can be used. To order to limit issues like overfitting and provide an overview into how the model is to extend to a different dataset etc. For k-fold cross-validation, the initial sample is divided randomly into k similar to k. The aim of the cross-validation is to create a data set to "test" the model in the training phase. One sub-sample of k is held for validation of the model, and the remaining k-1 sub-samples are used as train results. The method of cross-validation then is repeated k times (folds), and the validation data is used precisely once for each of the k subsamples. The outcomes of the folds can then be averaged in order to produce one estimate. The advantage of this approach is that all findings are used both for preparation and for confirmation, and that each measurement is used only once for validation. 10 fold cross-validation is often used but k remains an undefined parameter in general. The folds in stratified k fold are chosen to ensure that the average response value in all folds is approximately equal. Once defined as dichotomous, this implies that the ratio of each fold is approximately the same as the two class mark forms.

G. Performance Evaluation Metrics

Numerous performance evaluation tests have been used in this work to validate the performance of classifiers. We used a confusion matrix, each result is projected in precisely one chamber in the test array. There are two resting groups, 2×2 matrix. In comparison, there are two forms of accurate classifier estimation and two kinds of inaccurate prediction classifier. The following can be determined from the confusion matrix:

TP: expected performance as true positive (TP); we found an appropriate diagnosis of the HD topic, with cardiovascular disease.

We've noticed that a healthy person is properly labeled and the topic is safe. TN: The results predicted to be truly negative (TN).

FP: The conclusion is that it is wrong to classify a stable subject as having heart disease (Type 1 error): expected performance as false positive (FP).

FN: We consider that heart disease is incorrectly labeled as not having heart disease because the target is stable (type 2 error). Predicted to be false-negative (FN).

III. EXPERIMENTAL RESULTS AND DISCUSSION

Two separate diagnosis classes were included in this study: normal and patient who may have heart diseases. Many researchers have suggested different methods for the detection of heart disease as stated in the previous section. The accuracies recorded range from 50% to 87%. The collection includes 303 samples, 297 of which are full samples and six of which are incomplete samples. While the neural network ensemble model was trained in 70 percent of the cardiac disease database, the remaining 30% of the cardiac disease database was used for validating the proposed system. ROC curve is also one commonly used graph that summarizes the performance of a classifier over all possible threshold for assigning observations to a given class. Here, confusion matrix is used, which contains information about actual and predicted classifications performed by a classification system. The performance is evaluated using the data in the matrix. The results of the various machine learning algorithms are shown in figure 1 From fig 1,

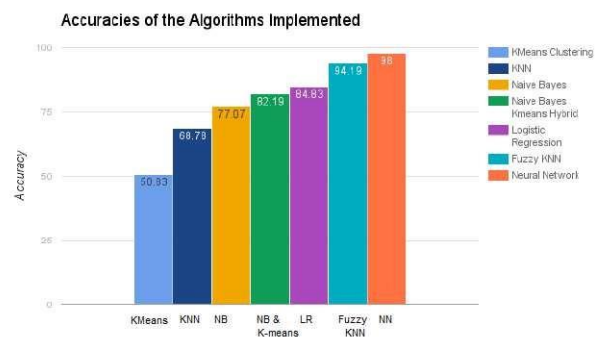


Fig. 1. Accuracies of various machine learning algorithms

It observed that the proposed system outperforms the existing methods of machine learning. Using a confusion matrix, the system classification findings were shown. The cell comprises the total number of copies listed as an appropriate combination of expected and actual network outputs in a confusion matrix. The following Table II display the confusion matrix showing the results of this network's classification.

Our reports were contrasted with previous findings recorded using different approaches. Table III demonstrates our system and previous approaches in terms of the classification accuracies.

ML Classifier	TP	TN	FP	FN
K Means Clustering	30.8	30.8	14.6	23.1
K Nearest Neighbors	31.2	39.5	14.1	15.2
Naive Bayes	33.1	43.5	11.1	12.3
K Means Clustering with Naive Bayes	35.1	47	7.7	10.2
Logistic Regression	37.7	47.9	6.9	7.5
Fuzzy K Nearest Neighbors	39.8	52.2	2.6	5.4
Neural Network	43.9	54.5	0.2	1.4

TABLE II CONFUSION MATRIX

Author/Year	Technique	Accuracy
Muhammad Fathurachman/2014 [14]	Extreme Learning Machine	84.00%
Divyansh Khanna, Rohan Sahu, Veeky Baths, Bharat Deshpande/2015 [15]	Logistic Regression	
	Generalized regression	82.80%
	Neural Network	89.00%
Kemal Polat, Seral Sahan, Salih Gunes/2005 [16]	FuzzyAIRSknn based system	87.00%
V Krishnaiah/2015 [17]	Fuzzy KNN	84.00%
	K Means Clustering with Naive Bayes	82.19%
Our Work	Fuzzy KNN	94.19%
	BP-Neural Network	98.00%

TABLE III

ACCURACY COMPARISON WITH LATEST RESEARCH PAPER

IV. CONCLUSION

In order to predict heart disease and then analyze the best method of analysis for treatment for the disease the works conducted a study of different algorithms, such as the Naive Bayes, the K Nearest neighbors, and logistic regression. Alongside simple algorithms, the hybrid algorithms generated are better evaluated and implemented than several previously cited research papers. From test results and evaluates, it is inferred that, compared to the success of K Means Clustering, KNN, logistic regression, etc., Neural Network and Fuzzy KNN appear to be better performance. In our approach we have introduced a regularization of the model which contributes to increasing the overall accuracy of the algorithm by overfitting. Fuzzy KNN promises that an incorrectly classified sample will not have a class membership next to one, whereas a correctly classified

sample has only one class membership. They are mindful of the complex nature of cardiac diseases. For this cause, our models still need professional advices and recommendations of implementation and the inclusion of some characteristics of patient information to determine the severity of heart disease, whether or not cardiovascular disease already occurs. In the future models of cardiac disease detection system can be further established. The research can be used to forecast such serious conditions in action and as an important tool for doctors and health professionals to counsel the patient appropriately. In the estimation of heart attack events, the model in this section will address more complex questions.

ACKNOWLEDGMENT

I am thankful to the Dr. Babasaheb Ambedkar Technological University, Lonere for the support and financial assistance received from under TEQIP-III program for my research work.

REFERENCES

- [1] P. A. Heidenreich, J. G. Trogon, O. A. Khavjou, J. Butler, K. Dracup, M. D. Ezekowitz, E. A. Finkelstein, Y. Hong, S. C. Johnston, A. Khera, *et al.*, "Forecasting the future of cardiovascular disease in the united states: a policy statement from the american heart association," *Circulation*, vol. 123, no. 8, pp. 933–944, 2011.
- [2] S. S. Yadav and S. M. Jadhav, "Machine learning algorithms for disease prediction using iot environment," *International Journal of Engineering and Advanced Technology*, vol. 8, no. 6, pp. 4303–4307., 2019.
- [3] S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, "Innovative artificial neural networks-based decision support system for heart diseases diagnosis," *Journal of Intelligent Learning Systems and Applications*, vol. 5, no. 3, p. 176, 2013.
- [4] J. Nahar, T. Imam, K. S. Tickle, and Y.-P. P. Chen, "Computational intelligence for heart disease diagnosis: A medical knowledge driven approach," *Expert Systems with Applications*, vol. 40, no. 1, pp. 96– 104, 2013.
- [5] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J.-J. Schmid, S. Sandhu, K. H. Guppy, S. Lee, and V. Froelicher, "International application of a new probability algorithm for the diagnosis of coronary artery disease," *The American journal of cardiology*, vol. 64, no. 5, pp. 304–310, 1989.
- [6] N. Cheung, *Machine learning techniques for medical analysis*. School of Information Technology and Electrical Engineering. PhD thesis, B. Sc. Thesis, University of Queensland, 2001.
- [7] S. Palaniappan and R. Awang, "Web-based heart disease decision support system using data mining classification modeling techniques," in *iiWAS*, pp. 157–167, 2007.
- [8] I. S. F. Dessai, "Intelligent heart disease prediction system using probabilistic neural network," *International Journal on Advanced Computer Theory and Engineering (IJACTE)*, vol. 2, no. 3, pp. 2319–2526, 2013.
- [9] F. Moreno-Seco, L. Mico', and J. Oncina, "A modification of the laesa algorithm for approximated k-nn classification," *Pattern Recognition Letters*, vol. 24, no. 1-3, pp. 47–53, 2003.

- [10] E. Alpaydin, "Voting over multiple condensed nearest neighbors," in *Lazy learning*, pp. 115–132, Springer, 1997.
- [11] J. Chen, Y. Xing, G. Xi, J. Chen, J. Yi, D. Zhao, and J. Wang, "A comparison of four data mining models: Bayes, neural network, svm and decision trees in identifying syndromes in coronary heart disease," in *International Symposium on Neural Networks*, pp. 1274–1279, Springer, 2007.
- [12] R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," *Expert systems with applications*, vol. 36, no. 4, pp. 7675–7680, 2009.
- [13] S. S. Yadav and S. M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *Journal of Big Data*, vol. 6, no. 1, p. 113, 2019.
- [14] M. Fathurachman, U. Kalsum, N. Safitri, and C. P. Utomo, "Heart disease diagnosis using extreme learning based neural networks," in *2014 International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA)*, pp. 23–27, IEEE, 2014.
- [15] D. Khanna, R. Sahu, V. Baths, and B. Deshpande, "Comparative study of classification techniques (svm, logistic regression and neural networks) to predict the prevalence of heart disease," *International Journal of Machine Learning and Computing*, vol. 5, no. 5, p. 414, 2015.
- [16] K. Polat, S. Sahin, and S. Güneş, "Automatic detection of heart disease using an artificial immune recognition system (airs) with fuzzy resource allocation mechanism and k-nn (nearest neighbour) based weighting preprocessing," *Expert Systems with Applications*, vol. 32, no. 2, pp. 625–631, 2007.
- [17] V. Krishnaiah, G. Narsimha, and N. S. Chandra, "Heart disease prediction system using data mining technique by fuzzy k-nn approach," in *Emerging ICT for Bridging the Future- Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1*, pp. 371–384, Springer, 2015.