Heart Disease Prediction Using Machine Learning

Budh Ram
B.Tech, CSE
IIITDM Jabalpur, India

Bhargavi Kishor Chetule
B.Tech, CSE
IIITDM Jabalpur, India

Kusum K. Bharti
Assistant Professor
IIITDM Jabalpur, India

---

1. Abstract :- Heart disease is the leading cause of death worldwide, with over 17.9 million deaths reported each year. Early detection and timely interventions can greatly reduce the morbidity and mortality associated with heart disease. Current approaches to heart disease prediction rely on subjective evaluation and may not be accurate or efficient enough to identify at-risk patients early. Therefore, there is a need for a more accurate and efficient approach to heart disease prediction.

Machine learning has emerged as a powerful tool for predicting heart disease, leveraging vast amounts of patient data to generate accurate and personalized risk assessments.However,current machine learning models for heart disease prediction suffer from limitations such as poor interpretability, lack of transparency, and the requirement for large amounts of training data. These limitations hinder the widespread adoption of machine learning models in clinical practice.

To address these limitations, The proposed work predicts the chances of Heart Disease and classifies the patient's risk level by implementing different Machine learning techniques. Thus, this paper presents a comparative study by analyzing the performance of different machine learning algorithms.

The outcomes of this study will have a significant impact on the medical community and individuals, providing an optimal approach to heart disease prediction and helping to reduce the morbidity and mortality associated with this devastating condition.

2. Introduction:

Heart disease is a major cause of morbidity and mortality worldwide, and early detection and intervention are critical for improving patient outcomes. Machine learning algorithms have shown promise in predicting the risk of heart disease by analyzing patient data, including demographic information, medical history, and laboratory test results.

The goal of heart disease prediction using machine learning algorithms is to develop accurate models that can identify patients who are at high risk of developing heart disease. These models can help healthcare providers to implement targeted interventions, such as lifestyle modifications or medication, to reduce the risk of developing heart disease.

Several machine learning algorithms have been used for heart disease prediction, including Logistic Regression, Naive Bayes Algorithm with Laplace Smoothing Technique, K-Nearest Neighbor Classification Technique, and Random forest. These algorithms work by analyzing patterns and relationships in patient data and using this information to predict the presence or absence of heart disease.

In recent years, there has been a growing interest in using machine learning algorithms for heart disease prediction due to their potential to improve patient outcomes and reduce healthcare costs. However, the effectiveness of these algorithms depends on several factors, including the quality and quantity of patient data, the choice of algorithm and feature selection technique, and the expertise of the healthcare provider in interpreting the results.

In this research paper, we aim to investigate the effectiveness of different machine learning algorithms for heart disease prediction and explore the importance of feature selection techniques in improving the accuracy of these algorithms. By doing so, we hope to contribute to the growing body of research on heart disease prediction using machine learning and provide insights into the potential benefits and limitations of these approaches.

3. Related work :

Lot of work has been carried out to predict heart disease using the UCI Machine Learning dataset. Different levels of accuracy have been attained using various Machine learning techniques ranging from traditional statistical methods to machine learning techniques. This literature survey aims to provide an overview of these approaches and methodologies and their effectiveness in predicting heart disease. These are explained as follows.

a. Traditional Statistical Method**:**-

Traditional statistical methods such as logistic regression, decision trees have been widely used for heart disease prediction. These methods work by analyzing the relationship between various risk factors and the likelihood of developing heart disease. However, these methods are limited in their ability to handle complex relationships and high-dimensional data.

b. Machine Learning Approaches**:**-

Machine learning approaches, on the other hand, have shown great potential in predicting heart disease. Some of the popular machine learning techniques used for heart disease prediction includes:-*Random Forest,Naive Bayes,*K-Nearest Neighbor Classification Technique etc.

1. The authors Nor Fatihah Zulkiflee and Mohd Saifullah Rusiman have proposed two methods for heart disease prediction, the LQD(Least Quartile Difference) method and the MAD(Median Absolute Deviation) method. The LQD method achieved an accuracy of 84.4% while the MAD method achieved an accuracy of 86.6%. These results show that both methods can effectively predict heart disease and that the MAD method outperforms the LQD method. However, more research is needed to validate the results and to compare these methods with other commonly used methods for heart disease prediction. It is important to note that these results may not generalize to all populations, as the study sample and factors used for prediction may differ from other studies.

2. The Authors Apurb Rajdhan, Milan Sai, Avi Agarwal, Dundigalla Ravi work proposed a paper which focuses mainly on various machine learning algorithms that are employed in heart disease prediction. It presents performance analysis of various ML techniques such as Naive Bayes, Decision Tree, Logistic Regression and Random Forest for predicting heart disease at an early stage. The result of this study indicates that the Random Forest algorithm is the most efficient algorithm with an accuracy score of 90.16% for prediction of heart disease. Lots of work has been carried out to predict heart disease using the UCI Machine Learning dataset.

3. Author Uma N Dulhare's (2018) study conducted an evaluation of the effectiveness and efficiency of a proposed model with PSO as feature selection increases the accuracy of the Naive Bayes to classify heart disease study on the accuracy of predicting heart disease using Naive Bayes and particle swarm optimization. Linear classifier as a Naive Bayes (NB) is relatively stable with respect to small variation or changes in training data. Particle Swarm Optimization (PSO) is an efficient evolutionary computation technique which selects the most optimum features which contribute more to the result which reduces the computation time and increases the accuracy. The study proposed the future development of a recommendation system for early prediction of heart disease diagnosis, as well as further research on the use of particle swarm optimization for feature selection on datasets with a large number of features.

4. The Authors Vincy Cherian, Bindu M.S(2018) work proposed a paper which has One notable improvement to Naive Bayesian classification is the use of Laplace smoothing techniques. This method can enhance the accuracy of the algorithm, leading to more

reliable results in predicting patients with heart disease. The system is expandable and can incorporate more records or attributes, making it more adaptable to different datasets.
There are two prediction options in the system, 13 attributes, and 6 attributes prediction. The 13 attribute prediction is more accurate, with 86% accuracy, making it a better choice for predicting patients with heart disease. The system can also incorporate other data mining techniques to improve its predictive power further.

5. The Authors Nida Khateeb,Muhammad Usman work proposed a paper to review all of the classifiers used and apply them to a dataset to determine accuracy based on different feature reduction methods. Domain knowledge is also taken into consideration in order to identify the most relevant features that are the actual predictors of heart disease, while eliminating more general features from the dataset. Preprocessing is applied to the dataset to prepare it for classification and to achieve better accuracy.
The results of the study show that the highest accuracy achieved 79.20% by utilizing the Re-sampling Weka option and KNN (IBK) classifier with 14 attributes.
Overall, the research highlights the importance of selecting appropriate feature reduction methods and classifiers for heart disease prediction. It also emphasizes the need for accurate domain knowledge and proper preprocessing of datasets to achieve the best results. The study provides valuable insights into the different approaches for heart disease prediction and their respective accuracies.

Comparison on Performance

S.N	Author Name	Technique used	Performance	Remark
1.	Nor Fatihah Zulkiflee1 , Mohd Saifullah Rusiman1* (2021)	Logistic Regression	84.4% (LQD method) 86.6% (MAD method)	---
2.	Apurb Rajdhan, Milan Sai, Avi Agarwal, Dundigalla Ravi (2020)	Using different Algorithms	Decision Tree: Accuracy - 81.97% Precision - 0.845 Recall - 0.823 F-measure - 0.835	
			Logistic Regression: Accuracy - 85.25% Precision - 0.857 Recall - 0.882	

			F-measure - 0.869	----
			Random Forest: Accuracy -90.16% Precision - 0.937 Recall - 0.882 F-measure - 0.909	
			Naive Bayes: Accuracy - 85.25% Precision - 0.837 Recall - 0.911 F-measure - 0.873	
3.	Uma N Dulhare (2018)	Naive Bayes and particle swarm optimization.	Accuracy - 87.91%	Particle swarm optimization (PSO), Naive Bayes classifier
4.	Vincy Cherian, Bindu M.S (2017)	Naïve Bayes Algorithm and Laplace Smoothing Technique	Sensitivity - 0.75 Specificity - 0.961 Precision - 0.947 Accuracy - 86%	Smoothing Technique, Naïve Bayes Algorithm
5.	Nida Khateeb, Muhammad Usman	K-Nearest Neighbor Classification Technique	Naïve Bayes : After Re - Sampling Accuracy: 66.66 % Smote Accuracy: 56.19 %	Multilayer Perceptron, Smote class distribution
			KNN : After ReSampling Accuracy - 79.20 % Smote Accuracy - 73.41%	
			Decision Tree : After ReSampling Accuracy - 76. 89 % Smote Accuracy - 66.31%	
			Bagging : After ReSampling Accuracy - 76.23% Smote Accuracy - 70.50 %	

4. Proposed Methodology:

a. **Data Preprocessing:-**Data Preprocessing is a crucial step in the machine learning workflow. It involves preparing and transforming the raw data into a format that is suitable for machine learning algorithms. Data pre-processing includes various techniques such as data cleaning, normalization, scaling, feature selection, and feature extraction. The quality of the data used in the machine learning models directly affects the accuracy and performance of the models.Data cleaning involves removing irrelevant or redundant data, correcting errors, and handling missing values in the dataset.Data preprocessing not only enhances the quality of the data but also reduces the complexity and computational cost of the machine learning models. By removing irrelevant data, the model has less data to process, which leads to faster training and better accuracy. Additionally, data preprocessing helps in reducing overfitting and underfitting, which are common problems in machine learning.In conclusion, it improves the accuracy and efficiency of the machine learning algorithms by transforming the raw data into a format that is suitable for analysis.Hence, proper data preprocessing is critical to the success of predicting heart disease.

b. **Feature Selection:-** Feature selection is the process of selecting a subset of features from the original dataset that have the most significant impact on the outcome. It helps to reduce the complexity of the machine learning models and enhances their predictive power. In the case of heart disease prediction, selecting the relevant features that are most strongly associated with the disease is essential.Feature selection can significantly improve the accuracy of heart disease prediction models.It helps to reduce the dimensionality of the dataset, enhance the accuracy of the models, and improve their interpretability. By selecting the optimal features, we can improve the accuracy and efficiency of heart disease diagnosis and contribute to better patient outcomes. It is essential to identify the most contributing factors or features that have the most significant impact on the outcome. This helps to improve the accuracy of the model, reduce its complexity, and make it more interpretable.

   There are several methods to perform feature selection, such as filter, wrapper, and embedded methods. These methods can help to identify the most relevant features based on their statistical significance, correlation with the outcome variable, or their impact on the model's performance.In some cases, domain expertise can also be used to select the most contributing factors. For instance, in a heart disease prediction model, factors such as age, blood pressure, cholesterol level, and smoking history are known to be strongly associated with

the disease. Therefore, including these features in the model can help to improve its accuracy.

In conclusion, choosing the most contributing factors is a critical step in building an accurate and interpretable machine learning model. By using appropriate feature selection techniques and domain knowledge, we can identify the most relevant features and improve the model's performance. This can lead to better predictions, better understanding of the underlying factors, and ultimately better outcomes for the application at hand.

PROPOSED MODEL:

1. Data sources and data set :

The heart disease data were obtained from the website of the UCI Machine Learning Repository and consists of 303 samples. One response variable and 14 independent variables were included where the response variable indicates that '0' and '1' are the results of the absence or presence of heart disease.

The predictor variables are x1(age), x2(gender), x3(chest pain type), x4(Resting Blood Pressure- in mmHg), x5(Serum Cholesterol- in mg/dl), x6(Fasting Blood Sugar->120 mg/dl), x7(Resting Electrocardiographic), x8(Maximum Heart Rate), x9 (Exercise-Induced Angina), x10(oldpeak), x11(slope), x12(number of major vessels) and x13(thalassemia).

While the dependent variable y(presence of heart disease). This study should be able to identify the significant factors that affect the presence of heart disease and to find the prediction of the probability of the presence of heart disease.

	Attribute Description	Distinct Values of Attribute
1.	Age- represent the age of a person	Multiple values between 29 & 71
2.	Sex- describe the gender of person (0- Female, 1-Male)	0,1
3.	CP- represents the severity of chest pain a patient is suffering.	0,1,2,3

4.	RestBP-It represents the patient's BP.	Multiple values between 94& 200
5.	Chol-It shows the cholesterol level of the patient.	Multiple values between 126 & 564
6.	FBS-It represents the fasting blood sugar in the patient.	0,1
7.	Resting ECG-It shows the result of ECG	0,1,2
8.	Heartbeat- shows the max heart beat of patient	Multiple values from 71 to 202
9.	Exercise Induced Angina- used to identify if there is an exercise induced angina. If yes=1 or else no=0	0,1
10.	OldPeak- describes a patient's depression level.	Multiple values between 0 to 6.2.
11.	Slope- describes patient condition during peak exercise. It is divided into three segments(Unsloping, Flat, Down sloping)	1,2,3
12.	CA- Result of fluoroscopy.	0,1,2,3
13.	Thal- test required for patients suffering from pain in chest or difficulty in breathing. There are 4 kinds of values which represent the Thallium test.	0,1,2,3
14.	Target-It is the final column of the dataset. It is a class or label Column. It represents the number of classes in the dataset. This dataset has binary classification i.e. two classes (0,1).In class "0" represent there is less possibility of heart disease whereas "1" represent high chances of heart disease. The value "0" Or "1" depends on other 13 attribute.	0,1

2. Classification:

logistic regression:

It is a classification algorithm mostly used for binary classification problems. In logistic regression instead of fitting a straight line or hyper plane, the logistic regression algorithm uses the logistic function to squeeze the output of a linear equation between 0 and 1. There are 13 independent variables which makes logistic regression good for classification.

Randomized search cross-validation is a technique used to tune hyperparameters in machine learning models. It involves randomly sampling hyperparameters from a predefined search space and evaluating their performance on a validation set using cross-validation. This process is repeated 5000 times in this model, and the hyperparameters ('C': 3.861021229056552, 'max_iter': 5000, 'penalty': 'none', 'solver': 'sag') that produce the best performance are chosen as the final model. Combining lo gistic regression with randomized search cross-validation allows for efficient tuning of hyperparameters while also leveraging the strengths of the logistic regression algorithm. The resulting model is often more accurate and robust than a logistic regression model with manually tuned hyperparameters.

Naive Bayes :

Naive Bayes algorithm is based on the Bayes rule[]. The independence between the attributes of the dataset is the main assumption and the most important in making a classification. It is easy and fast to predict and holds best when the assumption of independence holds. Bayes' theorem calculates the posterior probability of an event (A) given some prior probability of event B represented by P(A/B) as shown in equation : P(A|B) = (P(B|A)P(A)) / P(B).

One common issue with Naive Bayes is that if a feature never appears in the training data, the probability estimate for that feature will be zero. This can cause problems when trying to classify new data points that have that feature. One way to address this issue is to use Laplace smoothing, also known as add-one smoothing, which adds a small value (usually 1) to all the counts in the probability estimate. To implement Naive Bayes with Laplace smoothing, we first need to split the data into training and testing sets, just as with any other classification algorithm. Then we need to calculate the probabilities of each feature for each class in the training data. To do this, use the following formula:

P(feature_i|class_j) = (count(feature_i and class_j) + 1) / (count(class_j) + num_features)

Here, count(feature_i and class_j) is the number of times feature i appears in instances of class j in the training data, count(class_j) is the total number of instances of class j in the training data, and num_features is the total number of features in the data. Once we have calculated the probabilities for each feature for each class, we can use Bayes' theorem to calculate the probability of each class for a new data point. Specifically, we can use the following formula:

P(class_j|data_point) = P(class_j) * ∏ P(feature_i|class_j)Here, P(class_j) is the prior probability of class j, which is simply the proportion of instances in the training data that belong to class j. P(feature_i|class_j) is the probability of a feature i given class j, which we calculated earlier. Finally, the product ∏ P(feature_i|class_j) is taken over all the features in the data point. To classify a new data point, we simply calculate the probability of each class using the above formula, and choose the class with the highest probability as the predicted class.

KNN :

k-nearest neighbors algorithm to predict whether a patient has heart disease or not based on various features. The dataset is loaded into a pandas dataframe, and the features and labels are separated into two arrays. The data is then split into training and testing sets using the train_test_split function from the sklearn.model_selection module.

Next, the StandardScaler function from the sklearn.preprocessing module is used to standardize the data by scaling the features so that they have a mean of 0 and a standard deviation of 1. This helps to improve the performance of the k-nearest neighbors algorithm.

A for loop is used to iterate over values of k from 1 to 30, and for each value of k, a KNeighborsClassifier object is created and trained on the training data. The object is then used to predict the labels for the test data, and the mean error is calculated and stored in a list.

A KNeighborsClassifier object is then created with k set to the value that gave the minimum error. The object is trained on the training data and used to predict the labels for the test data.

Random Forest:

The methodology of this algorithm involves several steps to improve the accuracy of a Random Forest Classifier model on a heart disease dataset.

First, the dataset is loaded using the Pandas library. The variance inflation factor (VIF) is then calculated for each feature in the dataset using the statsmodels library. The VIF is a measure of multicollinearity between features, with values greater than 1 indicating some degree of correlation. To improve model performance, features with a VIF greater than a specified limit (in this case, 11) are removed from the dataset using the drop function.

Next, the dataset is split into training and testing sets using the train_test_split function from the sklearn library. The features in the training set are then scaled using the StandardScaler function, which transforms the data so that each feature has a mean of

0 and a standard deviation of 1.
The Random Forest Classifier model is then trained using the training set. To optimize the model, the code performs a loop over a range of 500 random states for the RandomForestClassifier. For each iteration, the model is trained and the accuracy is computed on the testing set. If the current accuracy is greater than the previous maximum accuracy, the current accuracy is stored as the new maximum accuracy and the random state that produced it is stored as the best_x variable.
After the loop, the best random state is used to train the final Random Forest Classifier model on the training set, and the accuracy of the model is computed on the testing set using the accuracy_score function from sklearn.metrics. Finally, the classification_report function is used to print a report that includes precision, recall, and F1 scores for each class in the testing set.

5. Experiments results:

1. Logistic regression

	precision	recall	f1-score	support
0	0.89	0.86	0.88	29
1	0.88	0.91	0.89	32
accuracy			0.89	61
macro avg	0.89	0.88	0.88	61
weighted avg	0.89	0.89	0.89	61

2. Naive Bayes

	precision	recall	f1-score	support
0	0.84	0.90	0.87	29
1	0.90	0.84	0.87	32
accuracy			0.87	61
macro avg	0.87	0.87	0.87	61
weighted avg	0.87	0.87	0.87	61

3. KNN

	precision	recall	f1-score	support
0	0.90	0.79	0.84	33
1	0.85	0.93	0.89	43
accuracy			0.87	76
macro avg	0.87	0.86	0.86	76
weighted avg	0.87	0.87	0.87	76

4. Random Forest

	precision	recall	f1-score	support
0	0.92	0.89	0.91	27
1	0.91	0.94	0.93	34
accuracy			0.92	61
macro avg	0.92	0.92	0.92	61
weighted avg	0.92	0.92	0.92	61

Algorithm	Accuracy	Remarks
Logistic regression	88.52%	randomized search with hyperparameter
Naive Bayes	86.88%	Laplace smoothing
KNN	86.84%	feature scaling
Random Forest	91.8%	Feature selection using variance inflation factor and scaling

6. Discussion:

- Logistic Regression:

This model has an accuracy of 89% and performs well in terms of precision, recall, and F1-score for both classes (0 and 1). It correctly predicts 86% of the negative cases and 91% of the positive cases. It is a simple and efficient model that is widely

used for binary classification problems.

- Naive Bayes:

This model also has an accuracy of 87% and performs well in terms of precision, recall, and F1-score for both classes (0 and 1). It correctly predicts 90% of the negative cases and 84% of the positive cases. Naive Bayes is a probabilistic model that is based on Bayes' theorem and is widely used for text classification problems.

- K-Nearest Neighbors (KNN):

This model has an accuracy of 87% and performs well in terms of precision, recall, and F1-score for both classes (0 and 1). It correctly predicts 79% of the negative cases and 93% of the positive cases. KNN is a non-parametric model that is based on distance metrics and is widely used for classification problems.

- Random Forest:

This model has the highest accuracy of 92% and performs well in terms of precision, recall, and F1-score for both classes (0 and 1). It correctly predicts 89% of the negative cases and 94% of the positive cases. Random Forest is an ensemble model that combines multiple decision trees and is widely used for classification problems.

We can see that all four models (logistic regression, Naive Bayes, KNN, and Random Forest) achieved high accuracy scores ranging from 0.87 to 0.92. Looking at the precision, recall, and F1-score values, we can see that all models performed well in terms of identifying patients with heart disease (label 1), with precision, recall, and F1-score values ranging from 0.85 to 0.91. Overall, the Random Forest model had the highest accuracy and F1-score values, indicating that it may be the best model for this classification task. However, it is important to note that different models may perform better or worse depending on the specific dataset and problem being addressed, so it is always recommended to try multiple models and evaluate their performance before making a final decision.

7. Conclusion and Future direction:

Our aim is to address the research gap by investigating the performance of several machine learning models in heart disease diagnosis and recommending the best optimal model for accurate and efficient diagnosis. The study will use a publicly available dataset to train and test the models and compare their performance using various performance metrics. The results of research will provide insights into the effectiveness of different machine learning algorithms in heart disease diagnosis and guide the selection of the best optimal model for clinical applications. Overall, the study

aims to improve the accuracy and performance of heart disease diagnosis and contribute to better patient outcomes.

Based on the analysis and experimentation conducted in this research paper, it can be concluded that machine learning algorithms can be effectively used for heart disease prediction. In particular, the algorithms Logistic Regression, Naive Bayes Algorithm with Laplace Smoothing Technique, K-Nearest Neighbor Classification Technique, and Random forest have shown promising results for predicting the presence of heart disease in patients.

Overall, these results suggest that machine learning algorithms can be valuable tools in predicting heart disease and providing early intervention for patients at risk. Further research could explore the use of other algorithms and feature selection techniques to improve the accuracy and efficiency of heart disease prediction.

One potential area of research is the use of deep learning algorithms for heart disease prediction. Deep learning algorithms are a type of machine learning algorithm that can automatically learn features from raw data and have shown promise in other areas of healthcare, such as medical imaging analysis.

Another potential area of future research is the integration of multiple data sources for heart disease prediction. In addition to demographic information, medical history, and laboratory test results, there are other sources of data that may be relevant for heart disease prediction, such as genetic data or wearable sensor data. By combining multiple data sources, it may be possible to develop more comprehensive and accurate models for heart disease prediction.

Additionally, there is a need for research on the interpretability of machine learning models for heart disease prediction. While machine learning algorithms can be effective at predicting the presence of heart disease, they can also be opaque and difficult to interpret for healthcare providers. Research in this area could explore methods for explaining the predictions of machine learning models and providing insights into the factors that contribute to the risk of heart disease.

Finally, future research could investigate the implementation and adoption of machine learning algorithms for heart disease prediction in clinical settings. While the potential benefits of these algorithms are clear, there are also challenges to implementing them in practice, such as integrating them into electronic health records and ensuring that healthcare providers understand how to interpret the results. By studying the implementation and adoption of these algorithms, researchers can help to ensure that they are effectively used to improve patient outcomes.

7. Reference

[1] Avinash Golande, Pavan Kumar T, ”Heart Disease Prediction Using Effective Machine Learning Techniques”, International Journal of Recent Technology and Engineering, Vol 8, pp.944-950,2019.
[2] T.Nagamani, S.Logeswari, B.Gomathy,” Heart Disease Prediction using Data Mining with Mapreduce Algorithm”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-3, January 2019.
[3] Fahd Saleh Alotaibi,” Implementation of Machine Learning Model to Predict Heart Failure Disease”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019.
[4] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin, ”Design And Implementation Heart Disease Prediction Using Naives Bayesian”, International Conference on Trends in Electronics and Information(ICOEI 2019).
[5] Theresa Princy R,J. Thomas,’Human heart Disease Prediction System using Data Mining Techniques’, International Conference on Circuit Power and Computing Technologies,Bangalore,2016.
[6] Nagaraj M Lutimath,Chethan C,Basavaraj S Pol.,’Prediction Of Heart Disease using Machine Learning’, International journal Of Recent Technology and Engineering,8,(2S10), pp 474-477, 2019.
[7] UCI, —Heart Disease Data Set.[Online]. Available (Accessed on May 1 2020): https://www.kaggle.com/ronitf/heart-disease-uci.
[8] Sayali Ambekar, Rashmi Phalnikar,“Disease Risk Prediction by Using Convolutional Neural Network”,2018 Fourth International Conference on Computing Communication Control and Automation.

[9] C. B. Rjeily, G. Badr, E. Hassani, A. H., and E. Andres, —Medical Data Mining for Heart Diseases and the Future of Sequential Mining in Medical Field,‖ in Machine Learning Paradigms, 2019, pp. 71–99.
[10] Jafar Alzubi, Anand Nayyar, Akshi Kumar. "Machine Learning from Theory to Algorithms: An Overview", Journal of Physics: Conference Series, 2018
[11] Fajr Ibrahem Alarsan., and Mamoon Younes ‘Analysis and classification of heart diseases using heartbeat features and machine learning algorithms’,Journal Of Big Data,2019;6:81.
[12] Internet source [Online].Available (Accessed on May 1 2020): http://acadpubl.eu/ap