International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST52411224

Detection of Cardiovascular Disease Using ECG Images in Machine Learning and Deep Learning

[1] Ms. K Jebima Jessy, [2]Dr G Manikandan, [3]Ms. S. Hemalatha, [4]Ms. Vilma Veronica

[1]PG Student, Kings Engineering College, Sriperumbudhur, Tamil Nadu, India
[2]Professor, Kings Engineering College, Sriperumbudhur, Tamil Nadu, India
[3]Assistant Professor, Kings Engineering College, Sriperumbudhur, Tamil Nadu, India
[4]Assistant Professor, Kings Engineering College, Sriperumbudhur, Tamil Nadu, India

ARTICLEINFO

Article History:

Accepted: 03 March 2024
Published: 13 March 2024

Publication Issue :
Volume 11, Issue 2
March-April-2024
Page Number :
131-139

ABSTRACT

One of the most prominent tools for detecting cardiovascular problems is the electrocardiogram (ECG). The electrocardiogram (ECG or EKG) is a diagnostic tool that is used to routinely assess the electrical and muscular functions of the heart. Even though it is a comparatively simple test to perform, the interpretation of the ECG charts requires considerable amounts of training. Till recently, the majority of ECG records were kept on paper. Thus, manually examining and re-examining the ECG paper records often can be a time-consuming and daunting process. If we digitize such paper ECG records, we can perform automated diagnosis and analysis. The main goal of this project is to use machine learning to convert ECG paper records into a 1-D signal. This can be achieved by extracting the P, QRS, and T waves that exist in ECG signals to demonstrate the electrical activity of the heart using various techniques. The techniques include splitting the original ECG report into 13 Leads, extracting and converting into the signal, smoothing, converting them to binary images using threshold and scaling. Post-feature-extraction, dimension reduction techniques like Principal Component Analysis are applied to understand the data. Multiple classifiers like k-nearest neighbors (KNN), Logistic Regression, Support Vector Machine (SVM), and Voting Based Ensemble Classifier are implemented, and based on the acceptable criteria on the accuracy, precision, recall, f1-score, and support, the model will be finalized. This final model will aid in the diagnosing of cardiac diseases, to detect whether a patient has/had Myocardial Infarction, Abnormal Heartbeat, or the patient is hale and healthy by inferring the ECG reports

Keywords : Electrocardiogram, Cardiovascular, K-Nearest Neighbors.

I. INTRODUCTION

According to the World Health Organization, heart disease is the first leading cause of death in the high and second leading cause of death in low-income countries. It has remained the leading cause of death at the global level for the last 20 years. This paper aims to analyze several data mining techniques implemented in recent years for diagnosing heart disease. At present, there are plenty of algorithms available that could detect and predict heart anomalies from clinical reports. However, in this project, the focus is more on discovering and extracting patterns from Electrocardiogram (ECG or EKG) image reports. By digitizing ECG records, the need for time consuming manual intervention for comprehending the report can be eliminated. With digitization, the automation of diagnosis and analysis can be achieved quicker. Many papers related to cardiovascular prediction focused on other features that included diet, age, gender, and many other dimensions, and then predicted for cardiovascular diseases based on these features. Our work is more on predicting diseases by providing the ECG chart to our model. Cardiovascular disease (CVD) refers to a class of diseases that involve the heart and blood vessels. It is a broad term that encompasses various conditions, including coronary artery disease, heart failure, stroke, and peripheral artery disease, among others. Cardiovascular diseases are among the leading causes of death worldwide.

The main underlying cause of cardiovascular disease is atherosclerosis, a condition characterized by the buildup of plaque within the arteries. Plaque consists of cholesterol, fat, calcium, and other substances, which gradually accumulate and narrow the arteries, restricting blood flow to the heart, brain, or other parts of the body. Several risk factors contribute to the development of cardiovascular disease, including:

- High blood pressure: Prolonged high blood pressure puts stress on the blood vessels and can lead to their damage.
- High cholesterol levels: Elevated levels of cholesterol in the blood can contribute to the formation of plaque in the arteries.
- Smoking: Tobacco use damages the blood vessels, increases blood pressure, and lowers good cholesterol (HDL).
- Diabetes: People with diabetes are at an increased risk of developing cardiovascular disease.
- Obesity: Excess weight puts strain on the heart and is associated with other risk factors such as high blood pressure and diabetes.
- Lack of physical activity: Sedentary lifestyles contribute to the risk of developing cardiovascular disease.
- Unhealthy diet: Diets high in saturated and trans fats, cholesterol, salt, and sugar can increase the risk of CVD.
- Family history: Having a close relative with cardiovascular disease increases the likelihood of developing it.
- Age and gender: The risk of cardiovascular disease increases with age, and men are generally at a higher risk than pre-menopausal women. However, the risk for women increases after menopause.
- Prevention and management of cardiovascular disease involve various approaches, including lifestyle modifications and medical interventions. These may include:
- Healthy lifestyle: Regular physical activity, maintaining a balanced diet, managing weight, quitting smoking, and limiting alcohol intake.

- Medications: Doctors may prescribe medications to control blood pressure, lower cholesterol levels, prevent blood clots, or manage other underlying conditions.
- Medical procedures: In certain cases, surgical interventions like angioplasty, bypass surgery, or stenting may be necessary to restore blood flow to the heart or other affected areas.
- Cardiac rehabilitation: This program involves supervised exercise, education, and counseling to help individuals recover from a heart attack, surgery, or manage their cardiovascular disease.
- Ongoing medical care: Regular check-ups, monitoring of blood pressure, cholesterol levels, and other risk factors are important in managing and preventing the progression of CVD.

It's worth noting that cardiovascular disease is a complex and multifactorial condition. If you have concerns about your cardiovascular health, it is important to consult with a healthcare professional for a personalized assessment and guidance.

II. EXISTING SYSTEM

This is the existing system and approach for cardiovascular disease prediction using ECG (electrocardiogram) images.

Here are a few commonly used methods:

- ✓ Convolutional Neural Networks (CNNs): CNNs have been successfully applied to ECG images for disease prediction. These networks are trained on large datasets of ECG images and can learn to extract relevant features automatically. CNNs can detect abnormalities in ECG patterns associated with various cardiovascular diseases.

III.PROPOSED SYSTEM

In the current era, systems are being developed for the automatic detection of cardiac-related issues. These systems predict high accuracy results based on one-dimensional ECG beats signals but are still not adopted as tools in health care institutes. The main areas that affect the success of these approaches, i.e., selection of feature, extraction techniques, types of classification algorithms, and the most important, the use of imbalanced data for classification can reduce the recognition accuracy of the minority class.

IV. LITERATURE REVIEW

[1] Title: Detection of Cardiovascular Diseases in ECG Images Using Machine Learning and Deep Learning Methods

Authors: Mohammed B. Abubaker; Bilal Babayiğit

Description:

Cardiovascular diseases (heart diseases) are the leading cause of death worldwide. The earlier they can be predicted and classified; the more lives can be saved. Electrocardiogram (ECG) is a common, inexpensive, and noninvasive tool for measuring the electrical activity of the heart and is used to detect cardiovascular disease. In this article, the power of deep learning techniques was used to predict the four major cardiac abnormalities: abnormal heartbeat, myocardial infarction, history of myocardial infarction, and normal person classes using the public ECG images dataset of cardiac patients. First, the transfer learning approach was investigated using the low-scale pretrained deep neural networks SqueezeNet and AlexNet. Second, a new convolutional neural network (CNN) architecture was proposed for cardiac abnormality prediction. Third, the aforementioned pretrained models and our proposed CNN model were used as feature extraction tools for traditional machine learning algorithms.

[2] Title: Detection of Cardiovascular Diseases in ECG Images Using Machine Learning and Deep Learning Methods

Authors: Mohammed B. Abubaker

Description:

Cardiovascular diseases (heart diseases) are the leading cause of death worldwide. The earlier they can be predicted and classified; the more lives can be saved.

Electro-cardiogram (ECG) is a common, inexpensive, and noninvasive tool for measuring the electrical activity of the heart and is used to detect cardiovascular disease. In this work, the power of deep learning techniques was used to predict the four major cardiac abnormalities: abnormal heartbeat, myocardial infarction, history of myocardial infarction, and normal person classes using the public ECG images dataset of cardiac patients. First, the transfer learning approach was investigated using the low-scale pre-trained deep neural networks SqueezeNet and AlexNet. Second, a new Convolutional Neural Network (CNN) architecture was proposed for cardiac abnormality prediction. Third, the afore-mentioned pretrained models and our proposed CNN model were used as feature extraction tools for traditional machine learning algorithms, namely Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), Decision Tree (DT), Random Forest (RF), and Naïve Bayes (NB). According to the experimental results, the performance metrics of the proposed CNN model outperform the exiting works; it achieves 98.23% accuracy, 98.22% recall, 98.31% precision, and 98.21% F1 score. Moreover, when the proposed CNN model is used for feature extraction, it achieves the best score of 99.79% using the NB algorithm.

[3] Title: Improving Disease Prediction by Machine Learning

Authors: Smriti Mukesh Singh1, Dr. Dinesh B. Hanchate2 – 2018

Description:

These days utilization of Big Data is expanding in biomedical and human services groups, exact investigation of medicinal information benefits early malady discovery, quiet care and group administrations. Fragmented therapeutic information lessens examination precision. The machine learning calculations are proposed for successful expectation of ceaseless infection. To beat the trouble of deficient information, Genetic algorithm will be utilized to remake the missing information. The dataset comprises of structured data and unstructured data. To extract features from unstructured data RNN algorithm will be utilized. Framework proposes SVM calculation and Naive Bayesian calculation for sickness expectation utilizing unstructured and structured information individually from hospital information. Community Question Answering (CQA) system is additionally proposed which will foresee the inquiry and answers and will give proper responses to the clients. For that, two calculations are proposed KNN and SVM. KNN algorithm will perform classification on answers and SVM calculation will perform classification on answers. It will help client to discover best inquiries and answers identified with infections.

[4] Title: Heart Disease Prediction System Using Data Mining Techniques

Authors: Abhishiek Taneja – 2015

Description:

In today's modern world cardiovascular disease is the most lethal one. This disease attacks a person so instantly that it hardly gets any time to get treated with. So diagnosing patients correctly on timely basis is the most challenging task for the medical fraternity. A wrong diagnosis by the hospital leads to earn a bad name and loosing reputation. At the same time treatment of the said disease is quite high and not affordable by most of the patients particularly in India. The purpose of this paper is to develop a cost-effective treatment using data mining technologies for facilitating data base decision support system. Almost all the hospitals use some hospital management system to manage healthcare in patients. Unfortunately, most of the systems rarely use the huge clinical data where vital information is hidden. As these systems create huge amount of data in varied forms but this data is seldom visited and remain untapped. So, in this direction lots of efforts are required to make intelligent decisions. The diagnosis of this disease using different features or symptoms is

a complex activity. In this paper using varied data mining technologies an attempt is made to assist in the diagnosis of the disease in question.

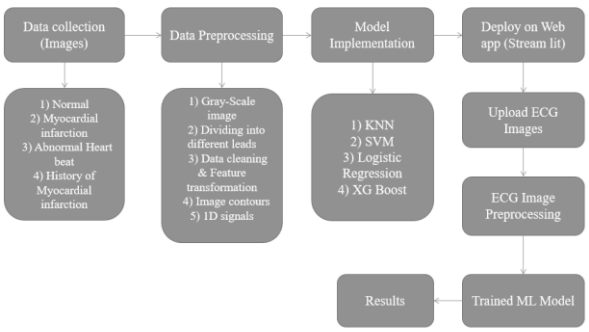
[5] Title: Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques

Authors: Chaitrali S. Dangare, Sulabha S. Apte - 2016

Description:

The Healthcare industry is generally "information rich", but unfortunately not all the data are mined which is required for discovering hidden patterns & effective decision making. Advanced data mining techniques are used to discover knowledge in database and for medical research, particularly in Heart disease prediction. This paper has analysed prediction systems for Heart disease using a greater number of input attributes. The system uses medical terms such as sex, blood pressure, cholesterol like 13 attributes to predict the likelihood of patient getting a Heart disease. Until now, 13 attributes are used for prediction. This research paper added two more attributes i.e., obesity and smoking. The data mining classification techniques, namely Decision Trees, Naive Bayes, and Neural Networks are analyzed on Heart disease database. The performance of these techniques is compared, based on accuracy. As per our results accuracy of Neural Networks, Decision Trees, and Naive Bayes are 100%, 99.62%, and 90.74% respectively. Our analysis shows that out of these three classification models Neural Networks predicts Heart disease with highest accuracy.

V. SYSTEM ARCHITECTURE



Module 1: Dataset Collection

A dataset (or data set) is a collection of data, usually presented in tabular form. Each column represents a particular variable. Each row corresponds to a given member of the dataset in question. It lists values for each of the variables, such as height and weight of an object. Each value is known as a datum.

- We have chosen to use a publicly-available Healthcare dataset which contains a relatively small number of inputs and cases. The data is arranged in such a way that will allow those trained in medical disciplines to easily draw parallels between familiar statistical and novel ML techniques. Additionally, the compact dataset enables short computational times on almost all modern computers. Dataset collection is the process of collecting, measuring and analyzing different types of information using a set of standard validated techniques. The main objective of data collection is to gather information-rich and reliable data, and analyze them to make critical business decisions. Once the data is collected, it goes through a rigorous process of data cleaning and data processing to make this data truly useful for businesses.

Here we have used 4 categories for image classification for our ECG images

- Normal
- Myocardial infarction
- Abnormal Heart beat
- History of Myocardial infarction

Module 2: Data Preprocessing

The sklearn.preprocessingpackage provides several common utility functions and transformer classes to change raw feature vectors into a representation that is more suitable for the downstream estimators. In general, learning algorithms benefit from standardization of the data set. If some outliers are present in the set, robust scalers or transformers are more appropriate. The behaviors of the different scalers, transformers, and normalizers on a dataset

containing marginal outliers are highlighted in Compare the effect of different scalers on data with outliers.

Standardization, or Mean removal and Variance Scaling

Standardization of datasets is a **common requirement for many machine learning estimators** implemented in scikit-learn; they might behave badly if the individual features do not more or less look like standard normally distributed data: Gaussian with **zero mean and unit variance.**

Scaling features to a range

In practice we often ignore the shape of the distribution and just transform the data to center it by removing the mean value of each feature, then scale it by dividing non-constant features by their standard deviation.

For instance, many elements used in the objective function of a learning algorithm (such as the RBF kernel of Support Vector Machines or the l1 and l2 regularizes of linear models) assume that all features are centered on zero and have variance in the same order. If a feature has a variance that is orders of magnitude larger than others, it might dominate the objective function and make the estimator unable to learn from other features correctly as expected.

An alternative standardization is scaling features to lie between a given minimum and maximum value, often between zero and one, or so that the maximum absolute value of each feature is scaled to unit size. This can be achieved using MinMaxScaler or MaxAbsScaler, respectively.

The motivation to use this scaling includes robustness to very small standard deviations of features and preserving zero entries in sparse data.

MaxAbsScaler works in a very similar fashion, but scales in a way that the training data lies within the range [-1,1] by dividing through the largest maximum value in each feature. It is meant for data that is already centered at zero or sparse data.

Normalization

Normalization is the process of **scaling individual samples to have unit norm**. This process can be useful if you plan to use a quadratic form such as the dot-product or any other kernel to quantify the similarity of any pair of samples.

This assumption is the base of the Vector Space Model often used in text classification and clustering contexts.

- Data preprocessing is an important step before inputting the data into the model for training. In this phase, data can be organized into meaningful information.
- Convert all ECG images to grayscale Resize image per requirements Divide the images into 12 sections for each image extracting 12 lead values (1-12) Remove Grid lines from each lead image and convert to binary image
- To trace and extract only the necessary signal from images, we used the contour technique. The image is converted to a one-dimensional signal. After that, MinMaxScaler is used to scale the image. For each lead(1-12) signal in all ecg image, save the 1D signal values in a.csv file. In a single csv file, combine all 12 lead values with the target label added.

Module 3: Model Implementation

- We tested various algorithms and then utilized the enseble technique to stack algorithms to enhance performance because it was a classification problem. Classification based on Clustering, KNN, Decision Trees. In this project, we perform different Supervised classification algorithms: k-nearest neighbors (KNN), Logistic Regression, Support Vector Machine (SVM), and

Voting Based Ensemble Classifier based on CSV DATA.

KNN Algorithm

- The K-Nearest Neighbors (KNN) algorithm is a popular supervised machine learning algorithm used for both classification and regression tasks. It is a non-parametric method that makes predictions based on the similarity of a new data point to its K nearest neighbors in the training set.

SVM

- Support Vector Machines (SVMs) are a type of supervised machine learning algorithm used for classification and regression tasks. They are particularly effective for binary classification problems but can also be extended to handle multi-class classification.
- The main idea behind SVMs is to find an optimal hyperplane that separates the data points of different classes in a high-dimensional feature space. A hyperplane is a decision boundary that separates the data into different classes based on their feature values. The optimal hyperplane is the one that maximizes the margin, which is the distance between the hyperplane and the closest data points of each class. SVMs aim to find this optimal hyperplane by solving an optimization problem.

Logistic Regression

- Logistic regression is a statistical algorithm used for binary classification tasks. It is a type of regression analysis that predicts the probability of an event occurring by fitting data to a logistic curve.
- In logistic regression, the dependent variable (the variable to be predicted) is binary, meaning it can take one of two possible values, such as "yes" or "no," "0" or "1," or "true" or "false." The independent variables (also called predictors or features) can be either continuous or categorical.

XG Boost

- XGBoost (Extreme Gradient Boosting) is a popular machine learning algorithm that belongs to the gradient boosting family. It is known for its efficiency and effectiveness in solving a wide range of machine learning problems, particularly in the field of structured data.
- XGBoost builds a predictive model by combining an ensemble of weak prediction models, typically decision trees, in a sequential manner. The algorithm initially fits a base model to the data and then iteratively improves it by adding more models that focus on the errors made by the previous models. This iterative process continues until a stopping criterion is met.

MODULE 4: Deployment of Web application

- By trained the machine learning algorithms SVM, KNN, Logistic Regression and XG Boost we generate the model file.
- By using the model file we implement the function in streamlit framework.

MODULE 5: Prediction

- In the web page we give input as ECG images, it will preprocess the data like grayscaling, contour, canny and 1d images.
- After preprocessing the images are converted to dimensionality reduction and after that It will predict and give output like which type of image Normal, Myocardial infarction, Abnormal Heart beat, History of Myocardial infarction.

VI. CONCLUSION

First, the user uploads ECG images to our web app. From there, image conversion techniques such as rgb2gray conversion, denoising, Gaussian Filtering, thresholding and contouring are implemented to extract signals without the grid lines. The signal is

then dimensionally reduced, and the necessary waves (P, QRS, T) are extracted using segmentation and fed into our pre-trained model from the analysis. Once the model finishes the analysis, it returns the results back to the user based on the findings. The empirical results show that we can produce faster and accurate predictions for heart patients by applying the given predictive model to the ECG images of new patients. This study can also be extended to include multiple different heart diseases if the feature extraction from images is done correctly and optimally along with increased accuracy of our modelHowever, it's important to note that ECG is not a standalone diagnostic tool. It should be used in conjunction with other clinical evaluations, medical history, and additional tests to arrive at a comprehensive diagnosis. Interpretation of ECG results requires expertise and experience, and the involvement of trained healthcare professionals is crucial for accurate interpretation and decision-making. In summary, ECG plays a vital role in the detection and assessment of cardiovascular conditions. Its ability to capture the electrical activity of the heart provides valuable information that aids in diagnosis, treatment planning, and ongoing monitoring of patients. By facilitating early detection and intervention, ECG contributes to better patient outcomes and improved cardiovascular health.

VII. REFERENCES

[1]. L. Wang, H. Zhang, K. C. Wong, H. Liu, and P. Shi, "Physiological-model-constrained noninvasive reconstruction of volumetric myocardial transmembrane potentials," IEEE Transactions on Bio-Medical Engineering, vol. 57, no. 2, pp. 296–315, 2019

[2]. Q. Zhang, D. Zhou, and X. Zeng, "HeartID: a multiresolution convolutional neural network for ECG-based biometric human identification in smart health applications," IEEE Access, vol. 5, pp. 11805–11816, 2017.

[3]. U. R. Acharya, S. L. Oh, Y. Hagiwara et al., "A deep convolutional neural network model to classify heartbeats," Computers in Biology and Medicine, vol. 89, pp. 389–396, 2017.

[4]. C. Potes, S. Parvaneh, A. Rahman, and B. Conroy, "Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds," in Proceedings of the 2016 Computing in Cardiology Conference (CinC), Vancouver, Canada, September 2016.

[5]. Baitharu, Tapas Ranjan, and Subhendu Kumar Pani. "Analysis of data mining techniques for healthcare decision support system using liver disorder dataset." Procedia Computer Science 85 (2016): 862-870.

[6]. Gavhane, Aditi, Gouthami Kokkula, Isha Pandya, and Kailas Devadkar. "Prediction of heart disease using machine learning." In 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 1275-1278. IEEE, 2018.

[7]. Singh, Smriti Mukesh, and Dinesh B. Hanchate. "Improving disease prediction by machine learning." Int J Res Eng Technol 5, no. 6 (2018): 1542-1548.

[8]. Taneja, Abhishek. "Heart disease prediction system using data mining techniques." Oriental Journal of Computer science and technology 6, no. 4 (2013): 457-466.

[9]. Dangare, Chaitrali S., and Sulabha S. Apte. "Improved study of heart disease prediction system using data mining classification techniques." International Journal of Computer Applications 47, no. 10 (2012): 44-48.

[10]. Thomas, J., and R. Theresa Princy. "Human heart disease prediction system using data mining techniques." In 2016 international conference on circuit, power and computing technologies (ICCPCT), pp. 1-5. IEEE, 2016.

[11]. Kaur, Beant, and Williamjeet Singh. "Review on heart disease prediction system using data mining techniques." International journal on recent and

innovation trends in computing and communication 2, no. 10 (2014): 3003-3008.

[12].Meyer, Alexander, Dina Zverinski, Boris Pfahringer, Jörg Kempfert, Titus Kuehne, Simon H. Sündermann, Christof Stamm, Thomas Hofmann, Volkmar Falk, and Carsten Eickhoff. "Machine learning for real-time prediction of complications in critical care: a retrospective study." The Lancet Respiratory Medicine 6, no. 12 (2018): 905-914.

[13].Rajkomar, Alvin, Michaela Hardt, Michael D. Howell, Greg Corrado, and Marshall H. Chin. "Ensuring fairness in machine learning to advance health equity." Annals of internal medicine 169, no. 12 (2018): 866-872.

[14].Rajamhoana, S. P., C. Akalya Devi, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika. "Analysis of neural networks based heart disease prediction system." In 2018 11th International Conference on Human System Interaction (HSI), pp. 233-239. IEEE, 2018.

[15].Ramalingam, V. V., Ayantan Dandapath, and M. Karthik Raja. "Heart disease prediction using machine learning techniques: a survey." International Journal of Engineering & Technology 7, no. 2.8 (2018): 684-687.

[16].Kohli, Pahulpreet Singh, and Shriya Arora. "Application of machine learning in disease prediction." In 2018 4th International conference on computing communication and automation (ICCCA), pp. 1-4. IEEE, 2018.

[17].Marimuthu, M., M. Abinaya, K. S. Hariesh, K. Madhankumar, and V. Pavithra. "A review on heart disease prediction using machine learning and data analytics approach." International Journal of Computer Applications 181, no. 18 (2018): 20-25.

[18].Beyene, Chala, and Pooja Kamat. "Survey on prediction and analysis the occurrence of heart disease using data mining techniques." International Journal of Pure and Applied Mathematics 118, no. 8 (2018): 165-174.

[19].Khourdifi, Youness, and Mohamed Bahaj. "Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization." International Journal of Intelligent Engineering & Systems 12, no. 1 (2019): 242-252.

Cite this article as :

Ms. K Jebima Jessy, Dr G Manikandan, Ms. S. Hemalatha, Ms. Vilma Veronica , "Detection Of Cardiovascular Disease Using ECG Images in Machine Learning and Deep Learning", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 11 Issue 2, pp. 131-139, March-April 2024. Available at doi : https://doi.org/10.32628/IJSRST52411224
Journal URL : https://ijsrst.com/IJSRST52411224