IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, Nov 2022

Forecasting Systems for Heart Disease Using Advanced Machine Learning Algorithms

M. Jayanthi Rao[1], P. Prasanthi[1], P. Suresh Patnaik[2], M. Divya[1], J. Sureshkumar[3] and M. Ramanaiah[3*]

[1]Department of Computer Science and Engineering, Aditya Institute of Technology and Management, Tekkali - 532201, India.
[2]Department of Physics, Aditya Institute of Technology and Management, Tekkali - 532201, India.
[3]Department of Chemistry, Aditya Institute of Technology and Management, Tekkali - 532201, India.

** Corresponding author: ramanaiahmalla4@gmail.com*

Abstract

Heart disease is one of the most significant reasons of mortality in the current days. Now days it is very complicated task to predict the cardiovascular disease in the area of clinical data analysis. All the prediction is done by using manual approach which is becoming time complexity for the end users to find out the abnormalities. Hence this motivated me to designthe proposed application in order to efficient heart disease prediction using machine learning (ML) algorithm. In current days ML has been shown to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry. This is mainly because of its usage in different areas especially in the medical field. In this proposed work, we propose a novel model which is specifically that aims at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease. This proposed application is trained by using several ML algorithms and then check the following factors such as accuracy, precision, recall and F1- Score.By conducting various experiments on several ML Algorithms by taking UCI dataset, we finally check which algorithm fits best for efficient heart disease prediction.

Keywords: Machine Learning, Cardiovascular Disease, Healthcare Industry, Medical Field. Classification & Regression Techniques.

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, Nov 2022

1. Introduction

In 2008, almost 17.3 million individuals kicked the bucket because of heart disease. Over 80% of passings in world are a result of Heart illness. WHO assessed by 2030, practically 23.6 million individuals will kick the bucket because of heart problems as written in[1]. Expectation by utilizing information mining methods gives us precise consequence of ailment. IHDPS (shrewd coronary illness expectation framework) can find and concentrate shrouded information related with coronary illness from a chronicled coronary illness database. It can answer complex questions for diagnosing coronary illness and in this manner help medicinal services experts and specialists to settle on clever clinical choices which customary choice emotionally supportive networks can't. In this paper investigation of different information mining methods given in tables which were utilized and supportive for clinical experts or specialists for precise coronary illness conclusion.

Clinical analysis is one kind of investigation in the field of examination for AI, incompletely onthe grounds that the information is moderately organized and marked, and all things considered, this will be where patients initially cooperate with working, handy computerized reasoning frameworks. This is noteworthy for two reasons. Initially, as far as genuine patient measurements, clinical picture investigation is a litmus test regarding whether man-made brainpower frameworks will really improve understanding results and endurance. Besides, it gives a tried to human-AI collaboration, of how responsive patients will be towards wellbeing modifying decisions being made, or helped by a non-human entertainer.

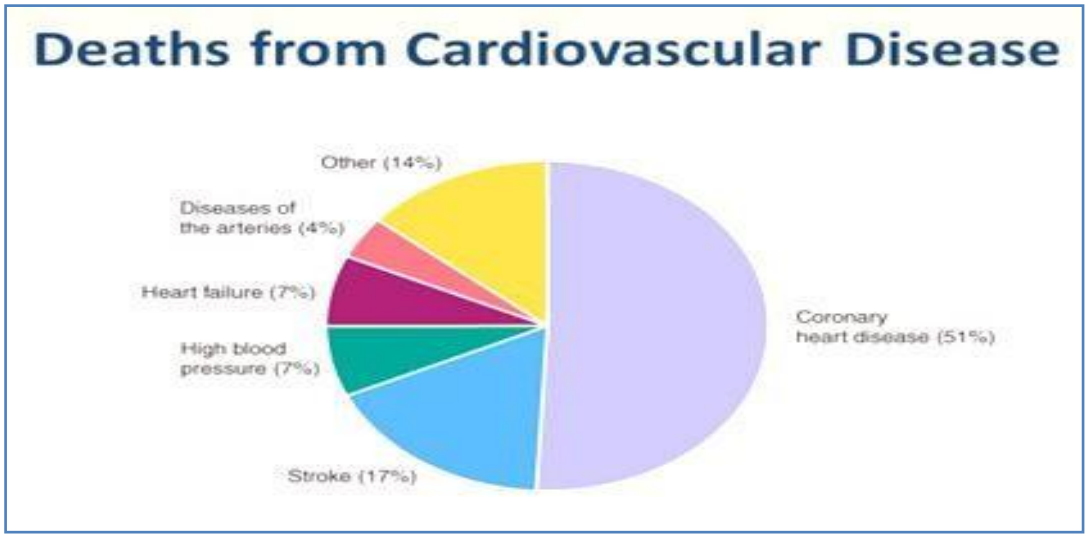


Figure 1. Represent the Major Cause of Deaths from Cardiovascular Disease

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, Nov 2022

From the above figure 1, we can clearly identify that some of the major causes of heart disease and its death which are occurred from cardiovascular diseases. The huge amounts of data generated for prediction of heart disease are too complex and voluminous to be processed and analyzed by traditional methods. Data mining provides the methodology and technology to transform these mounds of data into useful information for decision making. By using machine learning techniques, it takes less time for the prediction of the disease with more accuracy. So, in this present work we try to predict the heart diseases based on several machine learning classification algorithms and then check which algorithm gives best accuracy in terms of data retrieval and less time. Applying advanced machine learning techniques to heart disease treatment data can provide as reliable performance as that achieved in diagnosing heart disease.

2. Literature Survey

Literature survey is that the most vital step in software development process. Before developing the new application or model, it's necessary to work out the time factor, economy and company strength. Once all these factors are confirmed and got an approval then we can start building the application. The literature survey is one which is mainly deal with all the previous work which is done by several users and what are the advantages and limitations in those previous models. This literature survey is mainly used for identifying the list of resources to construct this proposed application.

2.1. Motivation

Santhana Krishnan J and Geetha S[2] Mohammed Abdul Khaleel[3] has given paper in the Survey of Techniques for mining of data on Medical Data for Finding Frequent Diseases locally. In this proposed work, the author mainly focuses on the distinct information about several data mining procedures and techniques which are required for medical knowledge processing. For example, heart infirmities, lung malignancy and lot more disease information‟s. The information mining or data mining is the process of extracting the valuable or useful information from hidden data and then try to analyze the data based on user inputs. Here the author mainly concentrated on the importance of Naïve Bayes Algorithm and its importance in order to classify the medical data. The used data- set is obtained from diabetic research institutes of Chennai, Tamilnadu which is leading institute. There are more than 500 patients in the dataset. The tool used is Weka and

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, Nov 2022

classification is executed by using 70% of Percentage Split. The accuracy offered by Naive Bayes is 86.419%.

Aditi Gavhane, Gouthami Kokkula, Isha Panday, Prof. Kailash Devadkar[4] Haik Kalantarian and Mohammad Pourhomayoun[5] have given a paper named Remote Health Monitoring Outcome Success prediction using First Month and Baseline Intervention Data. RHS systems are effective in saving costs and reducing illness. In this proposed work the authors mainly concentrated on the RHM framework, which is almost cell or mobile based for instructing the remote users and help to connect several users who wish to find the information related to medical. This RHM will help each and every individual to navigate the flow from one location to another and then gather the suitable information in very less time without any data loss.

L. Sathish Kumar and A. Padmapriya[6] has given a paper named Prediction for similarities of disease by using ID3 algorithm in television and mobile phone. In this proposed work the authors mainly concentrated on the major impact of coronary illness and how that is affecting the others. The given framework utilizes information mining methods, for example, ID3 algorithm[7]. This proposed method helps the people not only to know about the diseases but it can also help‟s to reduce the death rate and count of disease affected people. M.A.Nishara Banu and B.Gomathy[8] has given a paper named Disease Predicting system using data mining techniques. In this work the authors mainly concentrated on MAFIA (Maximal Frequent Item set algorithm) and K-Means clustering. As we all know that classification is major process for prediction of any disease present in humans, the authors concentrated on these two classification techniques and found which one is giving best accuracy and which one gives more efficient results. Wiharto and Hari Kusnanto have given a paper named Intelligence System for Diagnosis Level of Coronary Heart Disease with K-Star Algorithm[9]. In this paper they exhibit an expectation framework for heart infection utilizing Learning vector Quantization neural system calculation The neural system in this framework acknowledges 13 clinical includes as information and predicts that there is a nearness or nonattendance of coronary illness in the patient, alongside various execution measures. In this proposed work the authors concentrated on heart infection causing symptoms and various other factors which can increase the illness present in that human. The authors finally conclude that current diagnosis levels are not accurate in identifying the diseases and if the new way is proposed then we can achieve more efficient

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, Nov 2022

results.

3. Current System and Its Limitations

In the existing system there was no proper method to identify the heart disease prediction using data mining algorithms. The following are the main limitations in the existing system.

Limitation of Primitive System

1. More Time Delay in finding the root cause of heart diseases
2. There is no prevention technique due to late prediction.
3. There is no early prediction of heart disease.
4. There is no method to identify the heart diseases based on DM Method

4. Proposed System and Its Advantages

In the proposed system used several machine learning (ML) Classification algorithms[10-12] for efficient heart disease prediction. The proposed system uses ML-Approach for classifying the each and every factor very accurately for diagnosis of heart disease and then check which one is best for identifyingthe disease[13] in very accurate manner and less time complexity to retrieve the data.

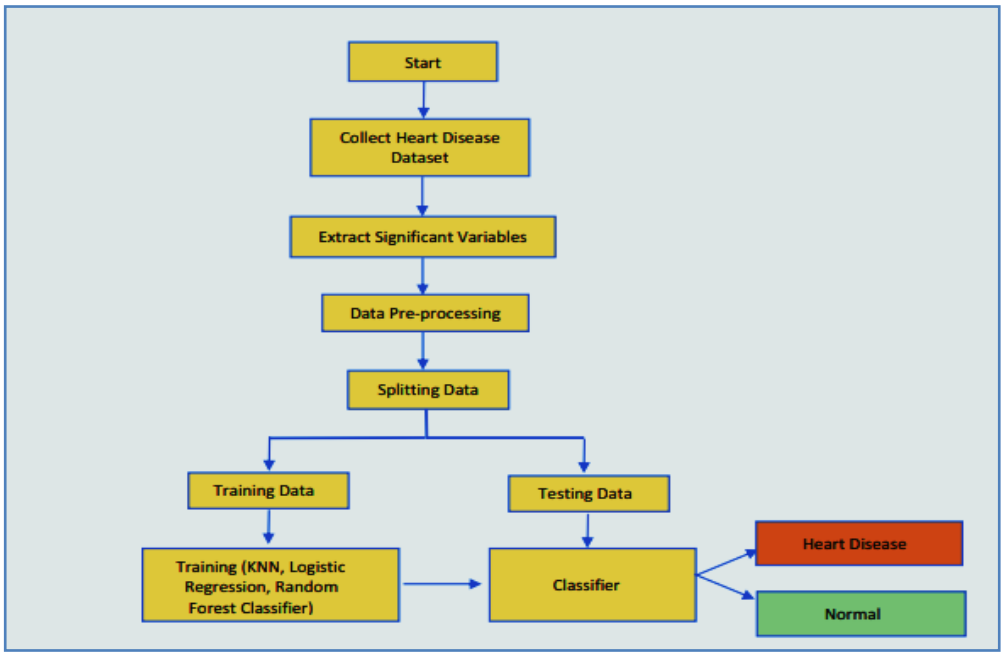


Figure 2. Represent the Flow of Proposed Model for Heart Disease Classification

Table.1 Attributes of the Dataset

S. No.	Attribute	Description	Type
1	Age	Patient's age (29 to 77)	Numaric
2	Sex	Gender of patient(male-0female-1)	Nominal
3	Cp	Chest pain type	Nominal
4	Trestbps	Resting blood pressure(in mm Hg on admission to hospital ,values from 94to 200)	Numerical
5	Chol	Serum cholesterol in mg/dl, values from 126 to564)	Numerical
6	Fbs	Fasting blood sugar>120mg/dl, true-1 false-0)	Nominal
7	Resting	Resting electrocardiographic sresult (0 to 1)	Nominal
8	Thali	Maximum heart rate achieved(71 to 202)	Numerical
9	Exang	Exercise includ edagina(1-yes 0-no)	Nominal
10	Oldpeak	ST depression introduced by exercise relative to rest(0 to .2)	Numerical
11	Slope	The slop of the peak exercise ST segment (0 to1)	Nominal
12	Ca	Number of major vessels(0-3)	Numerical
13	Thal	3-normal	Nominal
14	Targets	1 or 0	Nominal

Advantages of the Proposed System:

1. By using advanced machine learning techniques, it takes less time for the prediction of the disease with moreaccuracy.
2. In this paper we survey different papers in which one or more algorithms of advanced machine learning used for the prediction of heart disease.
3. Applying advanced machine learning techniques to heart disease treatment data can provide as reliableperformance as that achieved in diagnosing heart disease.

5. Implementation Phase

Implementation is the stage where the theoretical design is converted into programmatically manner. In this stage we will divide the application into a number of

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, Nov 2022

modules and then coded for deployment. The front end of the application takes Google Collaboratory and as a Back-End Data base we took UCI Heart Patients Records as dataset[14,15]. Here we are using Python as Programming Language to Implement the current application[16]. The application is divided mainly into following 5 modules. They are as follows:

1. Import Necessary Libraries
2. Load Dataset Module
3. Data Pre-Processing
4. Train the Model Using Several ML Algorithms
5. Find the Performance of ML Algorithms

Now let us discuss about each and every module in detail as follows:

5. 1. Import Necessary Libraries

In this module initially we need to import all the necessary libraries which are required for building the model. Here we try to use all the libraries which are used to convert the data into meaningful manner. Here the data is divided into numerical values which are easily identified by the system, hence we try to import numpy module and for plotting the data in graphs and charts we used matplot library.

Importing essential libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline

import os
print(os.listdir())

import warnings
warnings.filterwarnings('ignore')
```

```
['.ipynb_checkpoints', 'heart.csv', 'Heart_disease_prediction.ipynb', 'README.md']
```

5. 2. Load Dataset Module

In this module we try to load the dataset which is downloaded or collected from UCI repository. Here we store the dataset names as „Heart.csv „file and this dataset contains the following information such as:

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, Nov 2022

Data	Columns	Total Columns
Age	303	Non-null int64
Sex	303	Non-null int64
Cp	303	Non-null int64
Trestbps	303	Non-null int64
Chol	303	Non-null int64
Fbs	303	Non-null int64
Resting	303	Non-null int64
Thali	303	Non-null int64
Exang	303	Non-null int64
Oldpeak	303	Non-null float64
Slope	303	Non-null int64
Ca	303	Non-null int64
Thal	303	Non-null int64
Targets	303	Non-null int64
Dtypes	303	Float64(1), int64
Memory usage	303	33.3 kb

Table-2 Load Dataset Module

Every attribute contains some information which are tested and collected based on individual patient id.

5.3 Data Pre-Processing Module

Here in this section, we try to pre-process the input dataset and find out if there are any missing values or in-complete data present in the dataset. If there is any such data present in the dataset, the application will ignore those values and load only valid rows which have all the valid inputs.

Age	Sex	Cp	Trestbps	Chol	Fbs	Resting	Thali	Exang	Oldpeak	Slope	Ca	Thal	Targets
63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

Table-3 Pre-processing Dataset

5.4. Train the Model Using Several Ml Algorithms

Here we try to train the current model on given dataset using several ML classification algorithms and then try to find out which algorithms suits best in order to identify and classify the input dataset accurately and efficiently. Here we try to use following algorithms on input dataset such as:

1. Logistic Regression
2. Naïve Bayes
3. Support Vector Machine
4. K-Nearest Neighbors
5. Decision Tree
6. Random Forest

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, Nov 2022

7. XGBoost
8. Neural Networks

5.5 Performance Analysis Module

Here in this module we try to compare each and every classification algorithm on given input dataset and then try to find out which one suits best for finding the accurate results. Finally we will identify the best algorithm which give accurate results in very less time. Here we can see **Random Forest** gives more accurate result compared with other ML Algorithms.

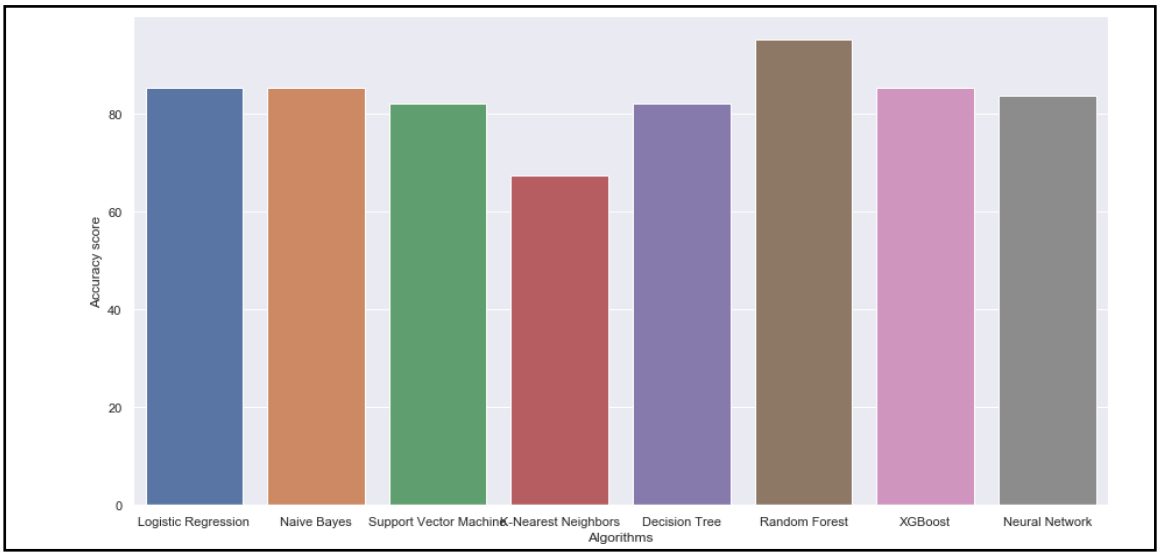
6. Experimental Results

In this section we try to design our current model using PYTHON as programming language and taking Heart Disease Dataset from UCI Machine Learning Repository as storage database. Here we try to construct the application by using several ML classification algorithms to predict heart disease present or not based on set of features which are recorded for every patient.

OUTPUT FINAL SCORE

In: scores =
[score_lr,score_nb,score_svm,score_knn,score_dt,score_rf,score_xgb,score_nn]
algorithms = ["Logistic Regression","Naive Bayes","Support Vector Machine","K-NearestNeighbors","Decision Tree","Random Forest","XGBoost","Neural Network"]
for i in range(len(algorithms)):
print("The accuracy score achieved using "+algorithms[i]+" is:
"+str(scores[i])+" %")The accuracy score achieved using Logistic Regression
is: 85.25 %
The accuracy score achieved using Naive Bayes is: 85.25 %
The accuracy score achieved using Support Vector Machine is:
81.97 %The accuracy score achieved using K-Nearest Neighbors
is: 67.21 % The accuracy score achieved using Decision Tree is:
81.97 %
The accuracy score achieved using Random Forest is: 95.08 %
The accuracy score achieved using XGBoost is: 85.25 %
The accuracy score achieved using Neural Network is: 83.61 %

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, Nov 2022



From the above Graph we can conclude that Random Forest Gives Best Accuracy when compared with many other ML Algorithms for Efficient Prediction of Heart Disease Patients Dataset.

7. Conclusion

In the current work, we have come to the conclusion that the model we have suggested can provide accurate findings in order to identify heart disease patients very precisely based on a set of factors that are carefully observed based on input data. heart disease prediction utilizing various advanced machine algorithms. Because the heart is such an important component of our bodies, life depends on its effective operation. Other human bodily organs like the kidney, brain, and heart will be impacted if the heart is not functioning properly. Heart disease is a condition that affects how the heart works. We look at various articles that use one or more advanced machine algorithms to predict cardiac disease. The outcome of employing neural networks in heart disease is almost 100%. for the forecast to produce accurate results when applying cutting-edge machine learning techniques. Advanced machine algorithms can produce equally valid results when applied to heart disease treatment data as those obtained when heart disease is first diagnosed.

References

1. C.S. Dangare, S. S. Apte, A data mining approach for prediction of heart disease using neural networks, international journal of computer engineering and technology, 2012.
2. J. S. Krishnan and S. Geetha, “Prediction of Heart Disease using Machine Learning Algorithms” 2019 1st International Conference on Innovations in Information and

Communication Technology (ICIICT), 2019.

3. Deeanna Kelley "Heart Disease: Causes, Prevention, and Current Research" in JCCC Honors Journal.
4. Aditi Gavhane, Gouthami Kokkula, Isha Panday, Prof. Kailash Devadkar, "Prediction of Heart Disease using Machine Learning", Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology(ICECA), 2018.-04
5. Senthil kumar mohan, chandrasegar thirumalai and Gautam Srivastva, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" IEEE Access 2019.
6. Ponrathi Athilingam, Bradlee Jenkins, Marcia Johansson, Miguel Labrador "A Mobile Health Intervention to Improve Self-Care in Patients With Heart Failure: Pilot Randomized Control Trial" in JMIR Cardio 2017, vol. 1, issue 2, pg no:1.
7. Talasu Kavya, Dr. Jayanthi Rao Madina, T. Ravi Kumar, Efficient Heart Disease Prediction System Using Several Machine Learning Classification Algorithms, Design Engineering, 8, 4544-4553, 2021.
8. DhafarHamed, Jwan K. Alwan, Mohamed Ibrahim, Mohammad B. Naeem "The Utilisation of Machine Learning Approaches for Medical Data Classification" in Annual Conference on New Trends in Information & Communications Technology Applications – march, 2017.
9. Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients Mai Shouman, Tim Turner, and Rob Stocker International Journal of Information and Education Technology, Vol. 2, No. 3, June 2012.
10. N. Aditya Sundar, P. Pushpa Latha, M. Rama Chandra,performance analysis of classification data mining techniques over heart diseases data base, international journal of engineering science and advanced technology, 2012.
11. J. Bala Bhaskara Rao, M. Jayanthi Rao, M. Srinivasa Rao, A.D.S. Saketh, T. Ravi Kumar, Optimizing the design of a fly wheel using machine Learning. Neuroquantology, 20(12), 533-254, 2022.
12. J. Bala Bhaskara Rao, M. Jayanthi Rao, S. Paparao, A.D.S. Saketh, P. Anjaneyulu, computational analysis of a knuckle jointand implementation of the generalized regression neural network, 20(12), 3260-3271|, 2022
13. Shadab Adam Pattekari and Asma Parveen, prediction system for heart disease using naïve bayes, International Journal of Advanced Computer and Mathematical

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, Nov 2022

Sciences, 2012.

14. M. Jayanthi Rao, R. Kiran Kumar, J. Harikiran, Method for follicle detection and ovarian classification in digital ultrasound images using geometrical features. Journal of Advanced Research in Dynamical and Control Systems, 11, 1249-1258, 2019.
15. M. Jayanthi Rao and R. Kiran Kumar, Follicle Detection in Digital Ultrasound Images Using BEMD and Adaptive Clustering Algorithms, Lecture Notes in Mechanical Engineering. 2020, pp. 651–659.
16. M. Balakrishna, M. Ramanaiah, B. Ramakrishna, M. Jayanthi Rao and R. Neeraja: Inductively Coupled Plasma-Mass Spectroscopy: Machine Learning Screening Technique for Trace Elemental Concentrations In Hemidesmus Indicus. November, 2022, 65(1): 4431-4445.