2023 1st DMIHER International Conference on Artificial Intelligence in Education and Industry 4.0 (IDICAIEI) | 979-8-3503-3842-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/IDICAIEI58380.2023.10407008

Automated Heart Disease Prediction System using Machine Learning Approaches

Jaya Srivastava
Department of InformationTechnology
ABES Engineering College
Ghaziabad, India
jaya.srivastava@abes.ac.in

Lakshya Bhargva
*Department of Information Technology, ABES Engineering College*Ghaziabad, India
lakshya.20b0133008@abes.ac.in

Dhirendra Jai Prakash Yadav
Department of Information Technology
ABES Engineering College
Ghaziabad, India
dhirendra.19b131156@abes.ac.in

Monica
Department of Information Technology, ABES Engineering College
Ghaziabad, India
monika@abes.ac.in

Pranshu Saxena
Department of Information Technology, ABES Engineering College
Ghaziabad, India
pranshusaxena@gmail.com

Puneet Kumar Aggarwal
Department of InformationTechnology
ABES Engineering College
Ghaziabad, India
puneetaggarwal7@gmail.com

***Abstract*— The healthcare industry collects large, complex data. This information is not "extracted" to find hidden facts for effective decision-making. Heart disease, a non-communicable disease, is difficult to diagnose since it requires a patient's medical history. A accurate and effective automated system can detect cardiac problems. Modern data mining may solve this. In healthcare, association rule mining, classification, and clustering are used to predict cardiac disease. An automated Heart Disease Prediction System was created using data mining algorithms like Decision Trees, Naive Bayes, and Neural Networks. Results show that each strategy has a distinct advantage in reaching mining goals. Age, sex, blood pressure, and blood sugar can indicate heart disease risk. It helps establish important knowledge like heart disease trends and linkages. This showed that the prediction algorithm can accurately anticipate heart attacks.**

***Index Terms*— Heart Disease, Medical Data Mining, Clustering, Classification, Machine learning.**

I. INTRODUCTION

Findings indicate that every approach has a special advantage in achieving the specified mining aims. It is possible to estimate the probability of individuals developing cardiovascular disease using medical characteristics like age, sex, blood pressure, and blood sugar. It allows for the identification of similarities and links across medical factors associated with cardiovascular disease, and another essential knowledge. The findings demonstrated the effectivenessof the de Machine learning as one of the best methods for testing that is based on training and testing. A specific branch of Artificial Intelligence (AI), a broad area of study where machines mimic human abilities, ismachine learning. Machine intelligence is the combination of several technologies. Machine learning systems, in contrast, are equipped to comprehend how to analyze and use data. developed a strategy for forecasting cardiac disease. Machine learning is one of the best methods for testing that is based on training and testing. A specific branch of Artificial Intelligence (AI),a broad area of study where machines mimic human abilities, is machine learning. Machine intelligence is the combination of several technologies. Machine learning systems, in contrast, are ready to comprehend how to analyze and use data.

We employed the supervised learning methods Decision Tree, Support Vector Machine, and in this paper. Regression with logit. k-nearest Neighbour is one unsupervised learning algorithm. In this study, we evaluate the accuracy of four different machine learningtechniques, compare supervised and unsupervised algorithms, and determine which algorithm is most effective in accurately predicting heart disease.

Section 1 of this research paper comprises an introduction regarding machine learning algorithms and factors for heart diseases. Section II defined the machine learning categorization. Section III comprised the literature survey. Section IV regarding the approach applied to theheart disease prediction system. Section V comprised. various supervised and unsupervised algorithms used inthis project. Section VI concisely defines the resultbased on the dataset and their study. And the preceding Section VII comprised conclusion with a minor assessment of the future scope of this research.

II. MACHINE LEARNING TECHNIQUES

One effective way is based on two ideas: testing and training, i.e., the system learns openly from experience and information, and tests should be applied to differentneeds in line with the required algorithm. There are mainly three types of Machine Learning Algorithmsillustrated in Fig. 1.

A. Supervised Learning

Training under sufficient supervision or training while an instructor is on hand is referred to as supervised learning. There is always a trained dataset while analyzing a batch of data since we have a trained datasetthat acts as the instructor for making predictions on the supplied data. The foundation of supervised learning is the "train me" principle.

B. Unsupervised Learning

Unsupervised learning is defined as learning without supervision, i.e., without an instructor giving instructions. Unsupervised learning independentlyexplores any given set of

data to find patterns and associations between them. The new dataset is thenclassified and assigned to one of the associations based on these associations. The foundation of unsupervised learning is the concept of "self-sufficiency." Consider the

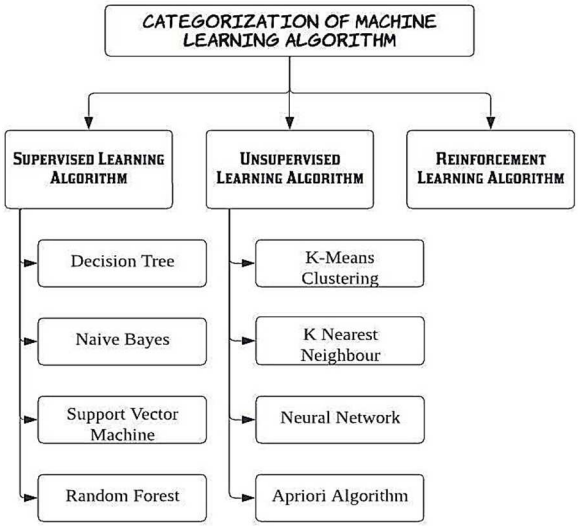

Fig.1. Categorization of Machine Learning Algorithm

scenario when the fruit has variations of the mango, banana, and apple. It would classify items into three unique categories using unsupervised learning based onhow they related to one another, and it would provide new information to specific groups on demand. Mango, banana, and apple are listed in supervised learning; however, three different groups were described in unsupervised learning. The technique of unsupervised algorithms is as follows:

- Dimensionality
- Clustering

There are following unsupervised machine learning algorithms (T-SNE, K-means clustering, PCA, k-nearest neighbours, Neural Network, and Apriori Algorithm).

C. Reinforcement

Reinforced learning refers to an entity's ability to interactwith its surroundings and determine the outcome. The "hit and trial" theory is the foundation of it. Each entity in reinforced learning has advantages and disadvantages and depending on the merit points, reinforced learning generates the dataset that was used to train the entity.

III. RELATED WORK

A few works have been done for heart disease prediction. Santosh Kumar Sundar [1] explains a simple model of machine learning for heart disease prediction by applying a hybrid approach using K-Means Algorithm and Naïve Bayes Algorithm. The intended model provides improved outcomes for forecasting the heart disease problem. it is an easy methodology to forecast the likelihood of heart diseases with simple patient health reports.

By using a genetic algorithm and fewer attributes, Anbarasi Masilamani and Anupriya [2] describe Enhanced Prediction of Heart Disease with Feature Subset Selection. The results demonstrate that the Decision Tree data mining strategy outperforms alternative data mining methods after incorporating feature subset selection with enhanced version model creation.

Benish Fida, Muhammad Nazir, Nawazish Naveed, Sheeraz Akram [3] explain Heart disease Classification Ensemble Optimization Using a Genetic Algorithm. A quite effective method for optimizing and finding high- quality solutions has been discovered with 98.63% accuracy in categorizing patients as having a disease or not comes from genetic algorithms.

Hnin Wint Khaing [4] uses the MAFIA (Maximal Frequent Itemset Technique) algorithm to describe Data Mining based Fragmentation and Prediction of Medical Data. The cardiac disease database is grouped using the K-means clustering algorithm. The results showed that the created forecasting system is capable of accurately forecasting a cardiac ailment with 84% accuracy, incontrast to a plain K-means MAFIA. Bhuvaneswari Amma N.G. [5] explains Cardiovascular Disease Prediction System using Genetic Algorithm and Neural Network using Genetic Algorithm founded aneural network for training the data The weighted base contains the neural network's final values, which are applied to estimate the heart disease risk. This method yielded a classification accuracy of 94.17%.

Dr. Priti Chandra, Dr. B.L. Deekshatulu, and M. Akhil Jabbar [6] explain how to predict heart disease using lazy associative classification. A lazy associative classification employed the data-centric contribution measure PCA to create class association rules for the detection of heart disease in Andhra Pradesh. Additionally, it provides some research findings that canhelp doctors make wise decisions.

M.A.Nishara Banu, and B.Gomathy [7] explain the Disease Forecasting System Using Data Mining Methods using K-means MAFIA with C 4.5 extension ofID 3 In comparison to MAFIA ID 3, the outcomes revealed that the proposed forecasting model can accurately diagnose heart disease with 89% success.

Richa Sharma [8] discusses the Efficient Heart Disease Prediction System Using a Decision Tree Using KEEL by Purushottam, Prof. (Dr.) Kanak Saxena, and Richa Sharma (Knowledge extraction based on evolutionary learning) Evolutionary algorithms for data mining issuesare evaluated using the free and open-source Java tool KEEL. The results show that the system has a large amount of promise for more accurate heart disease risk level prediction when the accuracy of the system isexamined in terms of classification accuracy.

Ali Miri, Salam Ismaeel, and Dharmendra Chourishi [9] Utilizing the Extreme Learning Machine (ELM) Method to Diagnose Heart Disease It resolves several problems that are difficult to manage in conventional techniques, such as local minimum, learning rate, halting criteria, and learning epochs. The ELM model with five outputs is more realistic than the backpropagation neuralnetwork (BPNN), which is less so. R. J. Thomas Using data mining techniques, Theresa Princy [10]

describes the Human Heart Disease Prediction System. The medical risk level is classified using data mining techniques as Nave Bayes, KNN, Decision Tree Algorithm, Neural Network, etc. The threat can be forecasted with great accuracy when more parameters are taken into account. Isha Pandya, Aditi Gavhane, and Gouthami Kokkula [11] explain Heart Disease Prediction. Create a programthat uses machine learning to predict the chance of getting a cardiac problem based on factors like age, sex,pulse rate, and other basic symptoms. The machine learning method with neural network models is one of the most reliable and accurate algorithms for predicting if a patient would acquire heart disease.

Heart Disease Prediction explained by Lakshmi CN, Bindhudhree M, Jaya Poojary, Manish C, and Shylaja B[12]. To calculate the accuracy score, supervised learning methods including Naive Bayes, decision trees, support vector machines, and random forest algorithms are used in a model. The accuracy score results of various classification methods for the training and testing set of data were mentioned using Python programming.

Heart disease forecast explained by authors [13] 14 significant features make up the data collection, which was used in conjunction with machine learning and deeplearning to finish the study. A lot of amazing results areproduced and evaluated using the accuracy and confusion matrix. Some unnecessary attributes in the dataset are managed using isolation forests, and data is also standardized to enhance results [13] [14]. Deeplearning produced an accuracy rate of 94.2%.

The study used the Multilayer Perceptron (MLP), Support Vector Machine (SVM), Random Forest (RF), and Naive Bayes classification algorithms to build the forecastingmodel (NB). Data pre-processing and feature selection stepswere conducted before building the models. The algorithmswere evaluated based on accuracy, precision, recall, and F1-score. The SVM model achieved a 91.67% accuracy rate [15].

All the works discussed above provide the accuracy of different machine learning algorithms that belong to the same type of learning either supervised or unsupervised. Some of them apply the genetic algorithmand neural network for feature selection and provide theaccuracy of the different machine learning algorithms. The work examined in this research paper is distinct as we compare the accuracy of Machine Learning algorithms of supervised learning and unsupervised learning.

IV. METHODOLOGY

The first step in the procedure of this system is data gathering, for that we utilize the dataset from the UCI repository that has been thoroughly checked by investigators and UCI authorities [16] [17].

A. Data Collection

The first phases in the forecasting procedure are data collection and dataset acquisition for training and testing.In this study, 36% of the data were utilized as the system's validation dataset, while 74% were applied to learning [18 [19].

B. Attribute Selection

Features of data are characteristics that are applied to systems, and for the human heart, numerous features are shown in TABLE.1 for prediction systems, such as age, sex, and heart rate.

TABLE. 1 ATTRIBUTE OF THE DATASET

S. No.	Attribute	Description
1	Age	Max Heart Rate
2	Sex	Patient Gender (male-0, female-1)
3	Resting Blood Sugar	Less than 100mg/dL (5.6 mmol/L) is normal
4	Resting ECG	Resting electrocardiography result (0 to 1)
5	Angina	Its type of chest pain caused by reduced blood flow to the heart
6	High ST segment	1: upward slope, 2: flat slope, 3: downward slope
7	Classification of Chest-pain	1 (typical), 2 (atypical), 3: pain, 4: asymptotic
8	Fasting Blood Sugar	Fasting blood sugar>120 mg/dl, true-1 false-0
9	Fluoroscopy colored vessels	Number of major vessels (0-3) colored by fluoroscopy
10	Thalassemia	Basically, haemoglobin carries oxygen 3-normal
11	Exercise induced ST	Exercise included Agina (1-yes 0-no)
12	Highest ST-segment	The slope of the peak exercise ST segment
13	Max Heart rate	71 to 202

D. Pre-processing of data

For machine learning algorithms to produce well-known results, pre-processing is necessary. For instance, we must manage zeros in the original raw data because the Random Forest approach does not support datasets with zeros. For our system, we must substitute fictional valuesfor some category values. In this process, we translate thenumber denoted by "0" and "1."

E. Data Balancing

Since the data leveling graph shows that each of the goal categories is equal, data leveling is necessary for precise results. The goal categories are shown in Fig. 2 as "0" heart disease patients and "1" patients lacking heart disease.

V. MACHINE LEARNING ALGORITHMS

A. Logistic regression

Logistic regression is a popular machine-learningalgorithm

that falls under the umbrella of supervised learning.

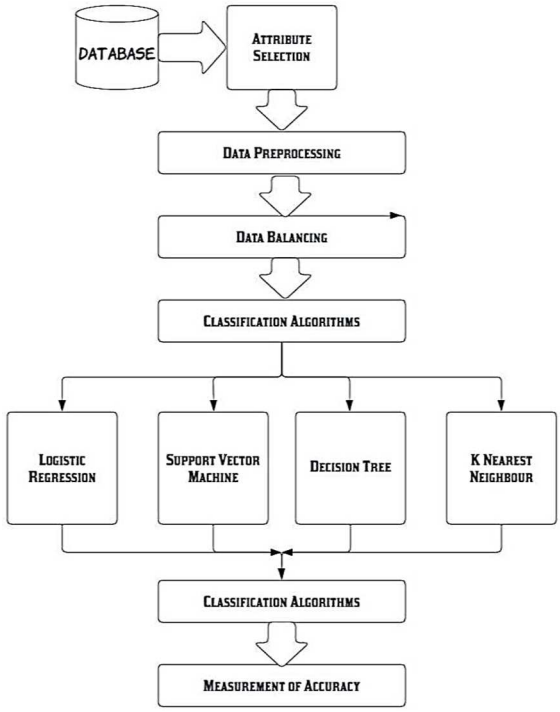

Fig. 2. Process flow of proposed methodology

When using logistic regression to categorize the findings using different types of datasets, the wide effective criteria for categorization can be easilydetermined [20] [21]. The logistic function is shown in Fig. 3. The sigmoid function, an arithmetic tool, is used to transform the projected values into possibilities. Any real value between 0 and 1 is changed into another value.The outcome of the logistic regression takes the shape ofan "S" curve and should be between 0 and 1, as it cannotgo above this value. The S-form curve is also known as the logistic function or sigmoid function. We apply the predefined margined value idea in logistic regression, which establishes the likelihood of either 0 or 1. Examples include numbers that incline to 1 over the margin value and to 0 under it.

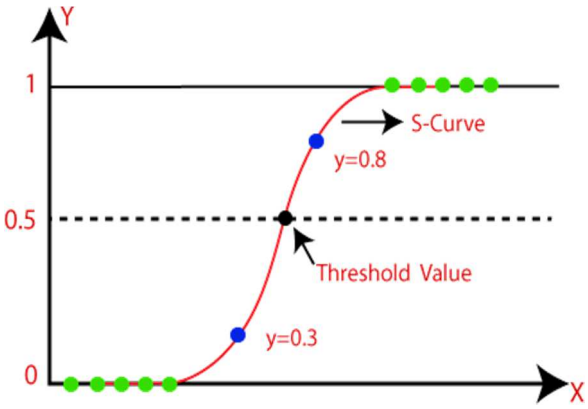

Fig. 3. Interpretation of sigmoid function [12]

B. Support Vector Machine

It is a type of machine learning methodology that relies on the idea of a hyperplane, that classifies the information by building a hyperplane among it. (Yi, Xi)is the instructional trial collection of data where $i = 1, 2, 3, \ldots, n$ and Xi is the i^{th} vector whereas Yi is the goal vector The quantity of hyperplanes determines the kind of support vector; for instance, if a straight line is employed as a hyperplane, the technique shown in Fig. 4. is known as a linear support vector [22][23].

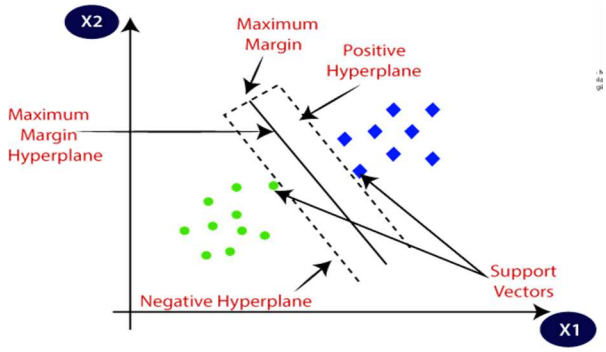

Fig. 4. Support Vector Machine [12]

C. Decision tree

Decision trees, on the other hand, are a class of supervised learning algorithms that graphically describedatasets. The root and other nodes of the tree are built depending on the probability of the data attributes, as shown in Figure 3. Eq. 1 is explaining the mathematical representation;

$$Entropy = -\sum P_{ij}.\log P_{ij} \quad (1)$$

D. K-nearest Neighbor

Based on the distance between the sites where the information is located, it groups various categories of information. The number of neighbours for each other's data sets is determined by the user, and this is a crucial element in the analysis of the dataset [26] [27] [28] [29].

The aforementioned Fig. 5. shows three neighbours, which denotes the presence of three different forms of data. The coordinates (Xi, Yi) are used to describe each group in two dimensions, where Xi is the x-axis and Yi is the y-axis and I is 1, $2, 3, \ldots, n$.

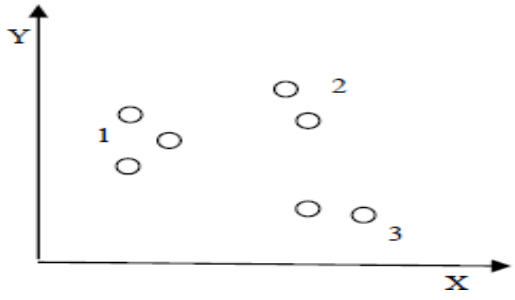

Fig. 5. K-nearest algorithm example

VI. RESULT ANALYSIS

This segment discusses the results for forecasting cardiac

illnesses using machine learning techniques for preparing and evaluating the data. This method's goal is improved accuracy, as well as excellent precision and recall metrics [24] [25]. The True-Positive (TP), True- Negative (TN), False-Negative (FN), and False-positive(FP) metrics can be created from these measurements.

$$Precision = \frac{T_P}{T_P + F_P} \times 100 \quad (2)$$

$$Recall = \frac{T_P}{T_P + F_N} \times 100 \quad (2)$$

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \times 100 \quad (4)$$

The statistical value of TP is defined as number of persons with heart diseases, TN is number of persons with heart diseases and no heart diseases, FP is number of persons with no heart diseases and finally FN isnumber of persons with no heart diseases and with heartdiseases.

A. Result of Logistic Regression

Performance matrices on the basis of different measure in presented in Fig. 6. 90% recall value is achieved.

	precision	recall	f1-score	support
0	0.85	0.71	0.77	41
1	0.79	0.90	0.84	50
accuracy			0.81	91
macro avg	0.82	0.80	0.81	91
weighted avg	0.82	0.81	0.81	91

Fig. 6. Performance matrices of Outcome

B. Result of Support Vector Machine

Different kernels have been tested for SVM classifiers. Among those classifiers linear is able to provide better precision.

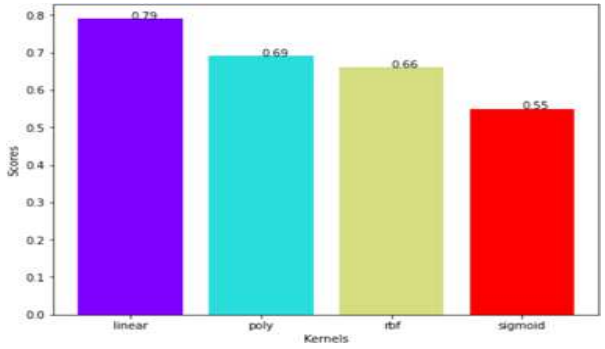

Fig. 7. Performance comparison of SVM on the basis of different kernels.

C. Result of Decision Tree

Classification results of decision tree is described in Fig. 8. Test is performed with different feature selected and after 11 features selected, algorithm is able to classify 75percent data sample correctly.

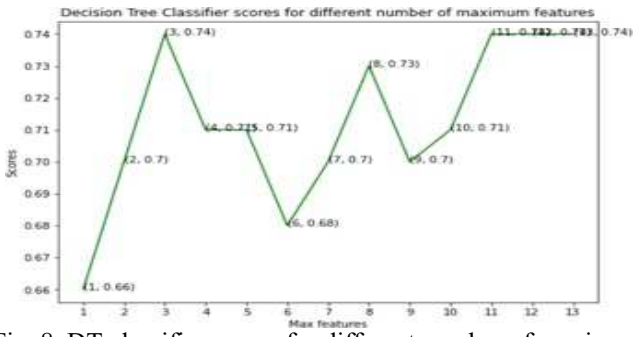

Fig. 8. DT classifier scores for different number of maximum features

Table 2. ACCURACY COMPARISON

Algorithm	Accuracy
Logistic Regression	83%
Support Vector machine	79%
Decision Tree	74%
K nearest Neighbour	64%

Accuracy should be calculated with the support of the confusion matrix of each algorithm with the number of counts of TP, TN, FP, and FN, and using equation (2), (3), and (4) of accuracy, the value has been calculated Fig. 8. Calculation of F1 score on the basis of differentfeatures selection

D. Results of K-Nearest Neighbor

	precision	recall	f1-score	support
0	0.62	0.49	0.55	41
1	0.64	0.76	0.70	50
accuracy			0.64	91
macro avg	0.63	0.62	0.62	91
weighted avg	0.64	0.64	0.63	91

Fig. 9. Performance analysis of knn algorithm

and it is concluded that the Logistic Regressionalgorithm of supervised learning has 83% accuracy which best among all algorithms and supervised learning is preferable rather than unsupervised learning algorithmK Nearest neighbour.

VII. CONCLUSION

Since the human heart is one of the body's key vital organs and heart disease forecasting is a major human issue, algorithm efficiency, and correctness is an important factor considered when evaluating an algorithm's efficiency. The dataset utilized for both training and evaluating purposes affects how accurate machine learning techniques are. We discover that logistic regression is the best approach when we compare them using the dataset whose attributes are presented in the Table 2.

REFERENCE

[1] Santhosh kumar Sundar, "A simple data mining model for heart disease prediction" , Journal of Global Pharma Technology 10(5):8-13, 2018.

[2] Shaik Akbar, P. Sri Silpa, Anand Thota, K. Nageswara Rao, "An Intangible System to Augment thePrediction of Heart Diseases Using Machine

Learning Techniques", International Journal of Recent Technology and Engineering, pp. 2277-3878,7(5), February 2019.

[3] Fida, Benish et al. "heart disease classificationensemble optimization using Genetic algorithm." *2011 IEEE 14th International Multitopic Conference* (2011): 19-24.

[4] Khaing, Hnin Wint. "Data mining-based fragmentation and prediction of medical data." *2011 3rd International Conference on Computer Research and Development* 2 (2011): 480-485.

[5] Amma, Narayanavadivoo Gopinathan Bhuvaneswari. "Cardiovascular disease predictionsystem using genetic algorithm and neural network." *2012 International Conference on Computing, Communication and Applications* (2012): 1-5.

[6] Jabbar, M. A., Deekshatulu, B. L., & Chandra, P. (2013). Heart Disease Prediction System using Associative Classification and Genetic Algorithm. ArXiv. https://doi.org/10.48550/arXiv.1303.5919.

[7] M.A.Nishara Banu and B. Gomathy "Disease Forecasting System using Data Mining Systems", International Conference on Intelligent ComputingSystems, 2014.

[8] Purushottam et al. "Efficient Heart Disease Prediction System." *Procedia Computer Science* 85 (2016): 962-969.

[9] Ismaeel, Salam et al. "Using the Extreme Learning Machine (ELM) technique for heart disease diagnosis." *2015 IEEE Canada International Humanitarian Technology Conference(IHTC2015)* (2015): 1-3.

[10] Mangesh Limbitote, Dnyaneshwari Mahajan, Kedar Damkondwar, Pushkar Patil, 2020, A Survey onPrediction Techniques of Heart Disease using Machine Learning, international journal of engineering research & technology (ijert) volume 09, issue 06 (june 2020).

[11] Lakshmi CN, Bindhudhree M, Jaya Poojary, Manish C, Shylaja B, "Heart Disease Prediction Using Machine Learning Algorithms" IJRASET, 10(4), 2022.

[12] Pranshu Saxena & Anjali Goyal, "Two-Stage Binary Classification of Follicular Histology Using Support Vector Machine", Chin. J. Med. Genet., vol. 31, no. 3, pp. 258–268, Jul. 2022.

[13] Pranshu Saxena & Anjali Goyal, "Computer- assisted grading of follicular lymphoma: a classification based on SVM, machine learning, and transfer learning approaches" The imagining Science Journal, Vol. 58, 2023.

[14] P. Jain, Anupam, P. K. Aggarwal, K. Makar, V. Shrivastava, S. Maitrey, "Machine Learning for Web Development: A Fusion", in Proceedings of AIST2020, 2020.

[15] Parita Jain, Amit Singhal, Diksha Chawla, Vineet Shrivastava, "Image Recognition and Segregation using Image Processing Techniques", TEST Engineering and Management, Vol. 83, pp. 2404 – 2410, 2020.

[16] Parita Jain, Puneet Kumar Aggarwal, Poorvi Chaudhary, Kshirja Makar, Jaya Mehta, and Riya Garg,"Convergence of IoT and CPS in Robotics", Emergenceof Cyber Physical Systems and IoT in Smart Automationand Robotics, pp.15-30, Springer, 2021

[17] Puneet Kumar Aggarwal, Parita Jain, Jaya Mehta, Riya Garg, Kshirja Makar, and Poorvi Chaudhary, "Machine Learning, Data Mining and Big Data Analytics for 5G-Enabled IoT", Blockchain for 5G enabled IoT: the new wave for Industrial Automation, pp. 351-375, Springer, 2021

[18] Jaskirat Singh, Puneet Kumar Aggarwal, and ParitaJain, "Property Rate Forecast Using Machine Learning", Design Engineering, Vol.7, pp. 9927-9949, 2021.

[19] Poorvi Chaudhary, Sachin Goel, Parita Jain, Mandeep Singh, Puneet Kumar Aggarwal, Anupam, "The Astounding Relationship: Middleware,Frameworks, and API", In Proceedings of the International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 1-4, 2021.

[20] Kshirja Makar, Sachin Goel, Prabhjot Kaur, Mandeep Singh, Parita Jain, Puneet Kumar Aggarwal, "Reliability of Mobile Applications: A Review and Some Perspectives", In Proceedings of the International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO),pp. 11-14, 2021.

[21] Puneet Kumar Aggarwal, Parita Jain, Poorvi Chaudhary, Riya Garg, Kshirja Makar, Jaya Mehta, "AIIoT for Development of Test Standards for Agricultural Technology", accepted for the book Intelligence of Things: AI-IoT Based Critical- Applications and Innovations, pp. 77-99, Springer, 2021.

[22] N. Gautam, A. Singh, K. Kumar, P K Aggarwal, Anupam, "Investigation on performance analysis of support vector machine for classification of abnormal regions in medical image", Journal of Ambient Intelligence and Humanized Computing, 2021.

[23] M. Singh, N. Sukhija, A. Sharma, M. Gupta, P. K. Aggarwal, "Security and Privacy Requirements for IoMT-Based Smart Healthcare System". Big Data Analysis for Green Computing, 17–37, Taylor &Francis, 2021.

[24] P. Jain, S. Sharma, Monica, P.K. Aggarwal, "Classifying Fake News Detection Using SVM, Naive Bayes and LSTM", in Proceedings of the International Conference on Cloud Computing, Data Science & Engineering- Confluence, 2022.

[25] V. K. Reshma, Ihtiram Raza Khan, M. Niranjanamurthy, Puneet Kumar Aggarwal, S. Hemalatha, Khalid K. Almuzaini, and Enoch Tetteh Amoatey, "Hybrid Block-Based Lightweight Machine Learning-Based Predictive Models for Quality Preserving in the Internet of Things- (IoT-) Based Medical Images with Diagnostic Applications", Computational Intelligence and Neuroscience, Vol. 2022, pp. 1-14, 2022.

[26] Parita Jain, Puneet Kumar Aggarwal, Kshirja Makar, Riya Garg, Jaya Mehta, Poorvi Chaudhary, "Machine Learning in Risk Analysis" Applications of Computational Science in Artificial Intelligence to be published by IGI Global, pp. 190-213, 2022

[27] Anupam Sharma, Mandeep Singh, Megha Gupta, Namrata Sukhija, Puneet Kumar Aggarwal, "IoT and Blockchain Technology in 5G Smart Healthcare", Blockchain Applications for Healthcare Informatics published by Elsevier, pp. 137-161, 2022.

[28] Pranshu Saxena, Kanu Priya, Sachin Goel, Puneet Kumar Aggarwal, Amit Sinha, Parita Jain, "Rice Varieties Classification Using Machine Learning Algorithms", Journal of Pharmaceutical Negative Results, Vol. 13, Issue 7, pp. 3762-3772, 2022.

[29] S. Goel, Monica, H. Khurana, P. Jain, "Social Media Analysis: A Tool for Popularity Prediction Using Machine Learning Classifiers", Design and Applicationsof Nature Inspired Optimization. Women in Engineeringand Science. Springer, Cham.