

Comprehensive Analysis of Atherosclerosis Disease Prediction using Machine Learning

Brajesh Kumar

PhD Scholar,
Department of CSE

Madhyanchal Professional University, Bhopal
brajeshpurbey2458@gmail.com

Dr. Harsh Mathur

Assistant Professor,
Department of CSE

Madhyanchal Professional University, Bhopal
harshmathur786@gmail.com

ABSTRACT

The prediction of atherosclerosis disease is a very complex process. The complexity of prediction faces a problem of multiple attributes of heart disease-related symptoms. The machine learning algorithm plays an essential role in the prediction of atherosclerosis disease. The various authors and research scholar proposed various algorithms based on machine learning and artificial intelligence. This paper presents the study of machine learning algorithms for the prediction of atherosclerosis disease. The analysis of atherosclerosis disease applied four machine learning algorithms such as support vector machine, KNN, decision tree and naïve Bayes. For the validation of algorithms applied four datasets: the applied dataset obtained from the UCI machine learning repository. For the evaluation of performance, measure three significant parameters like accuracy, specificity and sensitivity. The all-experimental work done in MATLAB environments. MATLAB is algorithm analysis software. The analysis of performance suggests that support vector machine is better than other machine learning algorithms.

KEYWORDS: - Heart Disease, Atherosclerosis, Machine Learning, UCI, MATLAB

INTRODUCTION

The human body is the integration of different organ, and each organ plays an essential role in the proper function of the human body[1]. The core organ of the human body is the heart and its proper functioning for the circulation of blood and oxygen all over the area of the body[2, 3]. In the current decade, the heart-related disease causes more mortality rate over the world reported by WHO. Some heart-related disease knows as heart attack, stroke, cardio arrest and coronary artery diseases. Coronary artery atherosclerosis is one of the leading causes of unexpected death all over the world[4]. The impact of atherosclerosis is well high and create Harding in the artery wall and reduces the flow of blood and oxygen in the human body[5, 6, 7]. Various factors cause atherosclerosis, such as blood pressure, diabetes, smoking, alcohol, gender, and history of the family most cases found in industrialized country over the world[8]. The early detection of such types of diseases saves human life. The detection process poses with automation and clinical strategy and measures cholesterol and thickness of blood the thickness and cholesterol case Harding and blockage in artery system[9, 10]. The medical organization is admitting the advancement of computer technology to automate the diagnosis process of atherosclerosis. The best way to treat atherosclerosis artery related disease change the lifestyle and early detection. Computer-aided diagnosis (CAD) plays a very significant role in heart-related disease. The various authors and research scholar applied machine learning methods for the detection of atherosclerosis disease. Machine learning provides supervised and unsupervised learning algorithm for the classification and grouping of behaviors of the disease. The machine learning algorithms such as support vector machine, decision tree naïve Bayes and some other algorithms enhance the prediction of heat-related illness and survive human life worldwide[11, 12, 13, 14]. The integration of machine learning algorithms with digital devices enables automatic detection of blood tests such as glucose pressure and some other parameters. In the current decade, various computer-aided methods are applied to detect atherosclerosis and heart-related disease[15]. Machine learning and feature-based methods enhance the process of detection[16, 17, 18]. The large scale of clinical data supports the analysis process of machine learning algorithms. Despite clinical risk analysis, the medical community has recognized the benefit of automatic discovery and prioritizing disease biomarkers. As a result, there has been significant work in interpreting machine learning approaches for ranking variables focused on finding ways to replace the relevance scores with measures that can be interpreted using standard methods[19, 20, 21, 22]. The feature selection-based methods applied the selection process of features is most common to impact heart disease[23, 24]. Various authors proposed feature-

based selection methods for the processing of features for classification. this paper focuses on the analysis of machine learning algorithms on a different dataset of atherosclerosis and heart-related diseases.

II. RELATED WORK

Accuracy of heart disease detection is very important factor, due to life threatening possibility if detection is not accurate. The process of incremental research various authors and research scholar contribute in algorithm modification and development of new algorithm for accurate detection. Some contribution of authors describe here.

Terrada, Oumaima Et al. [1] they fostered a novel AI MDSS to help the conclusion of cardiovascular sicknesses. Their investigation performed utilizing 835 patient clinical records that experience the ill effects of atherosclerosis, normally brought about by CAD, gathered from three data sets. The framework input layer incorporates a few info factors dependent on three information bases, the Cleveland coronary illness, Hungarian, and Z-Alizadeh Sani data sets. Seven autonomous classification strategies are applied to evaluate the framework: ANN, KNN, SVM, DT, NB, CE and DA calculations. The power of the examined strategies was assessed through a few execution measures. Terrada, Oumaima Et al. [2] atherosclerosis illness is the significant explanation that expands the death rate all throughout the planet. This disease finding gets troublesome because of the absentmindedness of beginning indications. Subsequently, there is a need to improve the expectation exactness of cardiovascular infections to limit treatment costs and to stay away from basic cases. This commitment is explained to characterize a MDSS of atherosclerosis illness. This framework can give coronary illness forecast utilizing the patient's clinical information. This MDSS depends on AI methods, for example, K-medoids and k-implies grouping for arrangement, ANN and KNN for expectation the presence and the shortfall of Atherosclerosis illness. The framework is approved on the Cleveland heart sicknesses data set. Yilmaz, Nihat Et al. [3] The main factors that forestall design acknowledgment from working quickly and adequately are the boisterous and conflicting information in data sets. This article presents another information arrangement technique dependent on bunching calculations for analysis of heart and diabetes illnesses. The informational indexes utilized in the determination of these illnesses are the Statlog (Heart), the SPECT pictures and the Pima Indians Diabetes informational indexes acquired from the UCI data set.

Bhatla, Nidhi Et al. [4] to calculation execution is tried in the coronary illness dataset which is taken from the UCI vault. The 14 credits are chosen from the dataset for the characterization. In particular, AI analysts have utilized Cleveland information base, especially consistently. The talked about work will likewise be contrasted and the current plan as far as precision, deficiency identification rate and execution time. Munger, Eric Et al. [5] They have fundamentally progressed their comprehension of bioprocesses in atherosclerosis, and in doing as such, they are starting to see the value in the intricacies, complexities, and heterogeneity atherosclerosis. They are additionally now better prepared to secure, store, and cycle the immense measure of organic information expected to reveal insight into the natural hardware included. Parameswari, C. Et al. [6] an examination idea is talked about zeroing in on a novel technique for order framework. This strategy is done with picture highlights got from fundus photos. It relies on the courses and vein grouping measure and furthermore by the morphological appearance.

Kolossváry, Márton Et al. [7] they describe the basics of radiomics, ML and DL, highlighting similarities, differences, limitations and potential pitfalls of these techniques. In addition, they provide a brief overview of recently published results on the applications of the aforementioned techniques for the non-invasive assessment of coronary atherosclerosis using CCTA. Singh, Navdeep Et al. [8] Most nations face high and developing paces of heart ailments or cardiovascular disease. Despite the way that, top tier drug is making the goliath proportion of information dependably, little has been done to utilize this open information to comprehend the difficulties that face a practical outline of echocardiography assessment occurs. To plan an insightful model for heart diseases affirmation utilizing information mining techniques that are good for improving the consistency of heart contaminations end. Learning Discovery in Database methodology including nine iterative and intuitive advances was gotten a handle on to think essential cases from a dataset a few echocardiography assessments report of heart patients over the globe. Rao, V. Sree Hari Et al. [9] they consider the clinical perceptions and propensities for people for anticipating the danger components of CHD. The ID of hazard factors helps in separating patients for additional serious tests like atomic imaging or coronary angiography. they present a novel methodology for foreseeing the danger elements of atherosclerosis with an in-fabricated ascription calculation and PSO. They contrast the presentation of their procedure and other AI strategies on STULONG dataset which depends on longitudinal investigation of moderately aged people going on for a very long time.

Qawqzeh, Yousef K. Et al. [10] A sample of 196 participants are enrolled in this study. Their CIMT test was recorded. The PPG's indices along with Age index are fed to a decision tree classifier developed in MATLAB to

predict and classify new data into high-risk atherosclerosis or normal atherosclerosis. The developed classifier showed promising results in which it revealed an overall accuracy of 82.6%. Additionally, it showed a sensitivity of 89.3% and specificity of 69.2%. These results represent a new possible method to be valid surrogate measure for atherosclerosis along with the used CIMT test. Animesh Hazra Et al [11] Heart illnesses are convoluted and remove bunches of lives each year. At the point when the early manifestations of heart sicknesses are overlooked, the patient may wind up with extreme results in a limited capacity to focus time. Stationary way of life and unnecessary pressure in this day and age have deteriorated the circumstance. On the off chance that the infection is identified early, it tends to be monitored. Be that as it may, it is consistently fitting to practice every day and dispose of undesirable propensities at the most punctual. Tobacco utilization and unfortunate weight control plans increment the odds of stroke and heart infections. Eating at any rate 5 helpings of leafy foods daily is a decent practice. For coronary illness patients, it is prudent to confine the admission of salt to one teaspoon each day.

Shouman, Mai Et al [12] This work shows the viability of an unaided learning procedure which is k-implies bunching in improving regulated learning strategy which is guileless bayes. It examines incorporating K-implies grouping with Naïve Bayes in the conclusion of coronary illness patients. It likewise researches various strategies for starting centroid determination of the K-implies grouping like reach, inlier, anomaly, irregular characteristic qualities, and arbitrary line techniques in the conclusion of coronary illness patients. Terrada, Oumaima Et al [13] This commitment is explained to characterize a MDSS of atherosclerosis infection. This framework can give coronary illness expectation utilizing the patient's clinical information. This MDSS depends on AI procedures, for example, K-medoids and k-implies bunching for arrangement, ANN and KNN for expectation the presence and the shortfall of Atherosclerosis illness. The framework is approved on the Cleveland heart infections information base. Han, Donghee Et al [14] RPP is related with occurrence cardiovascular occasions. Until now, no technique exists for the identification of people in danger of RPP at a solitary point on schedule. This examination coordinated coronary processed tomography angiography–decided subjective and quantitative plaque highlights inside a ML structure to decide its exhibition for anticipating RPP.

Nikan, Soodeh Et al [15] they talked about a calculation dependent on the AI strategies to foresee the danger of coronary supply route atherosclerosis. A REMI strategy is examined to gauge the missing qualities in the atherosclerosis information bases. A restrictive probability augmentation strategy is utilized to eliminate unessential traits and diminish the size of highlight space and in this manner improve the speed of the learning. The STULONG and UCI data sets are utilized to assess the examined calculation. Serrano, José Ignacio Et al [16] the information utilized was from the twenty years enduring essential preventive longitudinal investigation of the RF of atherosclerosis in moderately aged men. Study is named STULONG. The outcomes show that a few strategies foresee a few problems better than others, so it is fascinating to utilize every one of the calculations all at once and consider the outcome certainty dependent on the known propensity of every strategy. Georga, Eleni I. Et al [17] the important writing is broke down and stood out from regard to the procured dataset, the analyzed element space, the utilized prescient demonstrating plans and their discriminative or prescient limit CAD conclusion is presently performed by notable screening procedures, while CVD hazard can be evaluated by direct relapse models of gauge clinical, research center and anthropometric highlights, accepting linearity just as time-invariance of the basic info yield connections. Kumar, Abhishek Et al [18] The Data mining techniques are used to choose covered information that is important to medical care experts with compelling insightful dynamic. Information mining systems are used in the field of the medical care industry for various purposes. The target of this work is to evaluate and dissect utilizing three one of a kind information mining plan technique, for instance, NB, SVM and Decision Tree to choose the expected ways to deal with foresee the chance of coronary illness for diabetic patients reliant upon their prescient precision.

Magesh, G. Et al [19] they explore different avenues regarding their framework with Cleveland's heart tests from the UCI vault. Their group based CDTL essentially incorporates five key stages. From the outset, the first set has parceled through target mark conveyance. From the high dispersion tests, the other conceivable class blend has made. For each class-set blend, the critical highlights have recognized through entropy. With the huge basic highlights, an entropy-based parcel has made. Finally, on these entropy bunches, RF execution is made through critical and all highlights in the forecast of coronary illness. Rémy, Nfongourain Mougnotou Et al [20] This work gives a prescient model to choosing the most suitable medical care specialists, especially doctors, to analyze a patient. With regards to a multidisciplinary finding, this work gives an information mining model to recognize an expert doctor who can partake in such a determination and hence lessen the danger of blunders. To begin with, the model identifies the experts who can analyze a patient. The model uses the determined probabilities to give a positioning of expert doctors equipped for making a decent analysis. This positioning can be utilized to develop a gathering of experts who can partake in the multidisciplinary determination. Repaka, Anjan Nikhil Et al [21] Data mining, an extraordinary creating method that rotates around investigating and uncovering huge data from monstrous assortment of information which can be further valuable in analyzing and drawing out designs for settling on business related choices. Discussing the Medical area, execution of information mining in this field

can yield in finding and pulling out significant examples and data which can demonstrate helpful in performing clinical determination. The examination centers around coronary illness determination by thinking about past information and data.

Kazerouni, Faranak Et al [22] they applied four order models, including KNN, SVM, calculated relapse, and ANN for diagnosing T2DM, and they thought about the symptomatic force of these calculations with one another. they played out the calculations on six LncRNA factors and segment information. To choose the best presentation, they thought about the AUC, affectability, particularity, plotted the ROC bend, and showed the normal bend and reach. Shaji, Shaicy P. Et al [23] they are living in a postmodern period and there are gigantic changes happening to their day-by-day schedules which have an effect on their wellbeing decidedly and adversely. Because of these progressions different sort of illnesses is tremendously expanded. Particularly, coronary illness has become more normal nowadays. The existence of individuals is at a danger. Variety in Blood pressure, sugar, beat rate and so forth can prompt cardiovascular sicknesses that incorporate limited or obstructed veins.

Atallah, Rahma Et al [24] This work presents a larger part casting a ballot group technique that can foresee the conceivable presence of coronary illness in people. The forecast depends on straightforward moderate clinical trials led in any nearby center. Additionally, the point of this task is to give more certainty and precision to the Doctor's analysis since the model is prepared utilizing genuine information of solid and sick patients. The model groups the patient dependent on the greater part vote of a few AI models to give more exact arrangements than having just one model. At last, this methodology created an exactness of 90% dependent on the hard democratic troupe model. Bashir, Saba Et al [25] Heart infection is the vital justification short life. Enormous populace of individuals relies upon the medical care framework so they can get precise outcome in less time. Huge measure of information is created and gathered by the medical services association on the regular routine. To get captivating information, information advancement grants to separate the information through atomization of cycles. Weighted Association Rule is a kind of information mining method used to dispense with the manual undertaking which likewise helps in separating the information straightforwardly from the electronic records. This will help in diminishing the expense of administrations and furthermore helps in saving lives. Amin, Mohammad Shafenoor Et al [26] Cardiovascular infection is one of the greatest reason for grimness and mortality among the number of inhabitants on the planet. Forecast of cardiovascular infection is viewed as perhaps the main subject in the part of clinical information investigation. The measure of information in the medical care industry is immense. Information mining transforms the huge assortment of crude medical care information into data that can assist with settling on educated choice and expectation. There are some current examinations that applied information mining procedures in coronary illness forecast.

Dahiwade, Dhiraj Et al [27] Now-a-days, individuals face different illnesses because of the ecological condition and their living propensities. So the expectation of infection at prior stage becomes significant assignment. In any case, the precise expectation based on manifestations turns out to be excessively hard for specialist. The right expectation of sickness is the most difficult errand. To beat this difficult information mining assumes a significant part to foresee the infection. Clinical science has enormous measure of information development each year. Because of increment measure of information development in clinical and medical care field the exact examination on clinical information which has been profits by early tolerant consideration. Dhar, Sanchayita Et al [28] The point of this work is to introduce an effective method of foreseeing heart illnesses utilizing AI draws near. Consequently, they talked about a crossover approach for heart expectation utilizing Random woodland classifier and straightforward k-implies calculation AI methods. The dataset is additionally assessed utilizing two other distinctive AI calculations, in particular, J48 tree classifier and Naive Bayes classifier and results are analyzed. Results accomplished through Random backwoods classifier and the comparing disarray framework shows vigor of the approach. Dwivedi, Ashok Kumar Et al [29] Heart illnesses are of prominent general wellbeing dis-calm around the world. Heart patients are becoming rapidly inferable from deficient wellbeing mindfulness and terrible utilization ways of life. Subsequently, it is fundamental to have a structure that can viably perceive the pervasiveness of heart dis-ease in large number of tests quickly. At this crossroads, the capability of six AI methods was assessed for forecast of coronary illness. The presentation of these techniques was surveyed on eight assorted classification execution lists. Likewise, these techniques were evaluated on beneficiary usable trademark bend.

Rathnayakc, Bandarage Et al [30] This work presents an overview about various information mining and neural organization characterization advancements utilized in foreseeing the danger of happening heart sicknesses dependent on hazard factors. The critical level of an individual is ordering utilizing methods like K-Nearest Neighbor Algorithm, Decision Trees, Genetic calculation, Naïve Bayes and so forth and the precision is high when utilizing more credits and blends of above strategies. Goel, Sakshi Et al [31] With the expanding passing tallies because of heart infections, a framework is needed set up for precisely foreseeing coronary illness. Different specialists have examined various models for foreseeing heart illnesses utilizing various advances like fake neural

organizations, AI, information mining, and so on. This work examines the work done by different scientists on the exactness of coronary illness forecast through the various methodologies. A detail writing survey has been given in the examination. The investigation has likewise been introduced on the premise on innovation utilized. Haq, Amin Ul Et al [32] Heart illness is perhaps the most basic human infections on the planet and influences human existence severely. In coronary illness, the heart can't push the necessary measure of blood to different pieces of the body. Exact and on time determination of coronary illness is significant for cardiovascular breakdown avoidance and treatment. The determination of coronary illness through customary clinical history has been considered as not solid in numerous angles. To characterize the solid individuals and individuals with coronary illness, noninvasive-based techniques, for example, AI are dependable and productive.

Hasan, KM Zubair Et al [33] they talked about a novel classifier SDA - Sparse Discriminant Analysis technique for coronary illness identification. The time intricacy will be diminished in this calculation by ideal scoring investigation of LDA and will be complete to execute inadequate separation through the blend of Gaussians if limits between classes are nonlinear or if subgroups are accessible inside each class. Overall, contrasted with past procedures, their examined method is more proper for the analysis of coronary illness patients with higher exactness. Hossain, Rifat Et al [34] Obesity is an anatomical condition portrayed by a limit development of muscle to fat ratio. The corpulence rate is expanding progressively; from earlier examination, leanness is the genuine wellbeing infection in the globe. This examination gathered 259 information from determined metropolitan and country regions in regards to various danger factor of their everyday exercises. The motivation behind the investigation is to recreate the danger factor by utilizing SPSS, which assists with foreseeing the significant danger factor of stoutness by testing the class level characteristic as per cross-sectional examination with different qualities. Maji, Srabanti Et al [35] hybridization procedure is examined in which choice tree and artificial neural organization classifiers are hybridized for better execution of expectation of coronary illness. This is finished utilizing WEKA. To approve the presentation of the examined calculation, ten times approval test is performed on the dataset of coronary illness patients which is taken from UCI storehouse. The precision, affectability, and specificity of the individual classifier and half and half method are investigated. Maragatham, G. Et al [36] they researched whether utilization of profound figuring out how to show transient relations among occasions in EHRs would improve the model exhibition in anticipating introductory finding of HF contrasted with a portion of the conventional techniques that negligence transience. By inspecting these time stepped EHRs, they could perceive the relationship between different determination events lastly predicate when a patient is being investigated for an infection. Regardless, it is difficult to get to the current EHR information clearly, since practically all information is inadequate and not normalized.

Narayan, Subhashini Et al [37] A proficient clinical proposal framework has been examined in this work, specifically FTHDPS by utilizing Fourier change and AI procedure to anticipate the persistent heart illnesses adequately. Here, the info successions depend on the patient's time arrangement subtleties or information, which are disintegrated by Fourier change for removing the recurrence data. In FTHDPS, a sacking model is used for foreseeing the states of the patients ahead of time to create the total proposal. In FTHDPS, three classifiers are utilized, in particular fake neural organization, Naïve Bayes and backing vector machine, and genuine time arrangement constant coronary illness information are utilized to assess the examined model. The exploratory outcomes show that FTHDPS is much effective to give a solid and exact proposal to the heart patients. Rady, El-Houssainy A. Et al [38] Early recognition and portrayal are viewed as basic components in the administration and control of ongoing kidney sickness. Thus, utilization of proficient information mining methods is appeared to uncover and remove stowed away data from clinical and research center patient information, which can be useful to help doctors in boosting precision for distinguishing proof of illness seriousness stage. The aftereffects of applying PNN, MLP, SVM and RBF calculations have been looked at, and their discoveries show that the PNN calculation gives better grouping and expectation execution for deciding seriousness stage in ongoing kidney sickness. Raju, Cincy Et al [39] diagnosing patients accurately based on time is a urgent capacity for clinical help. An invalid conclusion done by the clinic leads for losing notoriety. The exact analysis of coronary illness is the prevailing biomedical issue. The inspiration of this work is to foster an effectual treatment utilizing information mining procedures that can help healing circumstances. Further information mining characterization calculations like choice trees, neural organizations, Bayesian classifiers, Support vector machines, Association Rule, K-closest neighbor order are utilized to analysis the heart sicknesses. Among these calculations SVM gives best outcome.

Ramasamy, S. Et al [40] In this work utilizing the affiliation rule digging calculation for extricate the coordinated with highlights from the medical clinic data set and watchword based bunching calculation is utilized to track down the precise sickness which is influenced by the patient. Both the calculations are utilized to acquire the exact outcomes with more proficiency and speedy handling. Raza, Khalid Et al [41] the three best performing arrangement calculations, in particular strategic relapse, multi-facet perceptron, and NB, and lion's share casting a ballot rule are applied to consolidate the yield of the classifiers. The model is powerful and solid as in a gathering of arrangement calculations is utilized for settling on the grouping choice and shielding it from erroneous order.

For a multivariate examination of information, having both discrete and consistent esteemed characteristics, the CHAID choice tree is developed for exploratory information investigation.

Safdar, Saima Et al [42] The momentum survey contributes with a broad outline of choice emotionally supportive networks in diagnosing heart illnesses in clinical settings. The specialists freely screened and preoccupied examinations identified with heart illnesses based clinical DSS distributed until 8-June-2015 in PubMed, CINAHL and Cochrane Library. The information extricated from the twenty full-text articles that met the consideration standards was ordered under the accompanying fields; heart sicknesses, techniques for informational collections development, AI calculations, AI based DSS, comparator types, result assessment and clinical ramifications of the announced DSS. Out of absolute of 331 examinations 20 met the incorporation measures. Sarkar, Bikash Kanti Et al [43] Heart illness is a main source of death on the planet. To drop its rate, viable and convenient finding of the sickness is extremely fundamental. Various mechanized choice emotionally supportive networks have been produced for this reason. In the current exploration, a prescient model comprising of two-level advancement is presented, to save lives and cost through powerful determination of the infection. Level-1 enhancement of the model initially recognizes parallelly an ideal extent for preparing and test sets for each dataset on equal machine. Singh, Poornima Et al [44] an EHDPS is created utilizing neural organization for anticipating the danger level of coronary illness. The framework utilizes 15 clinical boundaries like age, sex, circulatory strain, cholesterol, and stoutness for expectation. The EHDPS predicts the probability of patients getting coronary illness. It empowers critical information, e.g., connections between clinical components identified with coronary illness and examples, to be set up. They have utilized the multi-facet perceptron neural organization with backpropagation as the preparation calculation. The acquired outcomes have represented that the planned analytic framework can viably anticipate the danger level of heart sicknesses.

Woldemichael, Fikirte Girma Et al [45] Diabetes mellitus is fourth most high death rate infections on the planet and it is likewise a reason for kidney sickness, visual deficiency, and heart illnesses. Information mining strategies support a clinical choice for a right determination, treatment of illness in such way it limits the responsibility of subject matter experts. This investigation talked about to anticipate diabetes utilizing information mining strategies. Back spread calculation is utilized to anticipate if the individual has diabetic. And furthermore, J48, credulous bayes and backing vector machine were utilized to foresee diabetes. Palaniappan, Sellappan Et al [46] Studies have shown that heart illnesses have arisen as the main source of passing. Coronary illness is responsible for passing in all age gatherings and is basic among guys and females. A decent answer for this issue is to have the option to foresee what a patient's wellbeing status will resemble later on so the specialists can begin treatment a whole lot earlier which will yield better outcomes. It's much better than acting without a second to spare where the patient is as of now in danger and subsequently the forecast of coronary illness is generally investigated region. Yekkala, Indu Et al [47] the creators dealt with heart stalog dataset gathered from the UCI archive, utilized the Random Forest calculation and Feature Selection utilizing unpleasant sets to precisely foresee the event of coronary illness In this work, Random Forest classifier and Rough Set is utilized to foster a forecast model dependent on finding information to anticipate if the patient will experience the ill effects of coronary illness or not.

Hasan, S. M. M. Et al [48] by utilizing information acquire include choice method and eliminating superfluous highlights, distinctive grouping strategies with the end goal that KNN, ID3, Gaussian Naïve Bayes, Logistic Regression and Random Forest are utilized on coronary illness dataset for better expectation. Distinctive execution estimation factors like exactness, ROC bend, accuracy, review, affectability, explicitness, and F1-score are considered to decide the exhibition of the characterization strategies. Kolukisa, Burak Et al [49] they have evaluated a set of different classification algorithms, linear discriminant analysis and discussed a new hybrid feature selection methodology for the diagnosis of CHD. One of the advantages of the discussed method is its ability to work on real-time datasets. Throughout this research effort, they have tested the performance of their method using publicly available heart disease datasets. They have conducted comparative performance evaluations in terms of accuracy, sensitivity, specificity, F-measure, AUC and running time. Anitha, S., and N. Sridevi Et al [50] directed AI calculations specifically SVM, KNN and Naive Bayes are utilized to foresee the heart infections. The AI calculations are executed utilizing R programming language. The exhibitions of the calculations are estimated as far as exactness. The usefulness of the calculations is inspected and the results were thought.

Diwakar, Manoj Et al [51] they incorporate an audit of the classification techniques for AI and picture combination that have been shown to help medical care experts recognize coronary illness. They start with the AI brief and sum up depictions of the fundamentally utilized classification methods for diagnosing sicknesses of heart. At that point, they survey and show some work on the utilization of classification methods for AI and picture combination around here. It likewise gives an outline of the functioning calculation, and gives a portrayal of the current work. Malav, Amita Et al [52] The Heart Disease Prediction examined framework guides through a savvy choice

emotionally supportive network. In their talked about model a prescient examination is completed on UCI Heart Disease Data Set utilizing K-means and ANN information mining procedures. Clinical information is blend of fluffy and fresh qualities. Ismail, Ahmed Et al [53] The medical care information can be utilized to foster a wellbeing expectation framework that can improve in coronary illness anticipation. Huge information on medical care, including patient records, clinical notes, finding, guardians and family past illnesses, clinics, and sweep results can help in the period of infection recognizable proof and expectation. The arising AI strategy offers a significant system for estimating heart sicknesses. A high level SVM classifier was utilized by the program to direct boundary tuning to improve characterization exactness and execution. The examined work plans to foster a constant forecast framework for medical problems dependent on enormous clinical information handling on the cloud. In the examined adaptable framework, the clinical boundaries are shipped off Apache Spark to extricate the traits from the information and to apply the talked about AI calculation expecting to anticipate the medical services chances and send them as cautions and suggestions to the clients and the medical services suppliers too.

III. Methodology

Machine learning algorithms influence the automatic detection of heart related disease[3, 5]. The detection accuracy of algorithms is very challenging subject to study in case of heart related disease. The machine learning umbrella provides various algorithm for the detection of disease. In the process of analysis applied four machine learning algorithm such as support vector machine, decision tree, K-nearest neighbour and naïve bayes algorithms[7, 9, 10]. The support vector machine is most dominated classification algorithm for the analysis of medical clinical data diagnosis. Various authors modified and derived the support vector machine for the detection of heart related disease[12, 13]. Some others authors applied probability-based algorithm such as naïve bayes and decision tree algorithm[18]. The decision tree algorithm dominates the detection process based on the feature selection methods. some applied algorithms for the detection of heart related disease describe here.

SUPPORT VECTOR MACHINE (SVM)

SVM (Support vector machine) is machine learning algorithm derived by Vipin in 1990[10]. The support vector machine applied in various filed of image classification and pattern recognition. The nature of support vector machine is linear, non-linear and sigmoid. The non-linear support vector machine mapping the feature data with respect to one plane to another plan [2, 3, 16]. The separation of data plan is non-linear and decision factor correlate with margin function of support vector. The hyperplane of equation is derived as

$$WD.xi + b \geq 1 \text{ if } yi = 1 \quad (10)$$

$$WD.xi + b \leq -1 \text{ if } yi = -1$$

Here W is weight vector, x is input vector yi label o class and b is bias.

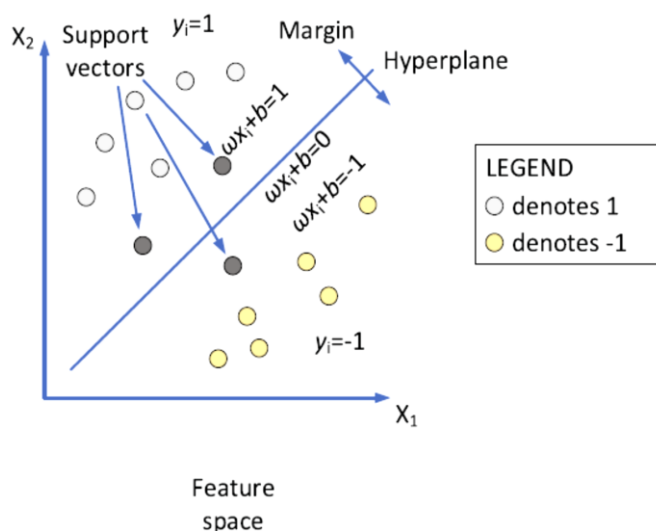


Figure 1: Process block diagram of support vector machine.

The minimization formulation of support vector

$$\begin{aligned} & \text{Minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \varepsilon_i, i = 1, 2, \dots, n \\ & \text{subject to } y_i(w^T D \cdot x_i + b) \geq 1 - \varepsilon_i \\ & \varepsilon_i \geq 0 \quad i = 1, 2, \dots, n \end{aligned} \quad (1)$$

Here C is constant, n is number of observation and ε_1 is slack variable.

The rule of decision function is

$$f(x) = \sum_{i=1}^n y_i a_i K(x_i, x_j) + b \quad (2)$$

DECISION TREE (DT)

A decision tree is predictive model of data classification. The processing of decision tree algorithm follows the tree structure of feature set[11, 29]. The partition of branch applied the concept of maximum entropy. The feature attribute has maximum entropy treat as leaf node class. The learning process of decision tree algorithm is very simple and suitable for the prediction of medical data analysis[45]. The decision tree rule formation according to the behaviours of feature attribute narrated in terms of maximum occurrence of probability. The core algorithm of decision tree is C4.5. the processing of these algorithm on design of tree in manner of top to down as class label. The entropy and gain are two functions manage the processing of decision tree algorithm[45, 53].

NAIVE BAYES (NB)

Naive bayes classifier work on the principle of feature correlation with class sample and other feature belong to other classes[13]. The processing of naïve algorithms describes in manners of opposite existence of features. The dimension of data is very high directly applied to naïve bayes algorithm for the classification. the naïve bayes algorithm suitable for the classification of medical clinical dataset[25, 26]. The naïve bayes algorithm describe here as

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \quad (3)$$

The processing of data based on normal distribution and the selection of real valued attribute. The other function cannot support the distributions of data. the naïve bayes algorithm estimate the joint probability $P(x, y)$. the probability of both x and y occurring given as at same time. The processing of attribute independent of each other's

$$P(x|y) = P(x) \times P(y) \quad (4)$$

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right) \quad (5)$$

K NEAREST NEIGHBOR CLASSIFIER (KNN)

The KNN classifier is simple algorithm of machine learning, it also knows as lazy classifier. The classification accuracy of KNN classifier varies in range of 70-80%. The major utility of KNN classifier in case of pattern recognition [25]. The KNN classification algorithm applied on the case of continuous nature of attribute. The processing of KNN algorithm describe here

1. Estimate K training attribute which belong to unknown attribute
2. Chose the common occurring classification of K

For the estimation of similarity in class of K instance applied different distance equation. The very famous distance equation is Euclidean distance equation.

IV. EXPERIMENTAL ANALYSIS

To evaluate the performance of different classification algorithm with MATLAB software. The version of software is R2014a, and the configuration of system is I7 processor, 16GB RAM and windows10 operating

system. The MATLAB provides the basic support library file of support vector machine and other classification algorithm. But the other function of classifier defines and programmed with function file and compile with library file. For the process of detection applied UCI machine learning dataset. The process of sample of data applied 10 cross folds for the processing of prediction and measurement of parameters such as accuracy, sensitivity and specificity.

DESCRIPTION OF DATASET : To evaluate the performance of machine learning algorithms, use various dataset of heart disease prediction. The resource of dataset is UCI machine Learning Repository. All dataset free available for study purpose. The description of dataset mention below.

- Hungarian: The Hungarian dataset has been collected at the Hungarian Institute of Cardiology, Budapest, by Andras Janosi. This database contains ten features. Through the 294 dataset samples, 34 samples were discarded because of missing values, and 262 records were commonly used and segregated in 62.21% healthy subjects and 37.78% with heart disease.
- Cleveland: This database contains 76 attributes, but consider only of 14 of them. The total number of instances is 303.
- Z-Alizadeh Sani: This dataset contains 270 instance and 13 attributes. Each patient could be in two possible categories CAD or Normal. A patient is categorized as CAD, if his/her diameter narrowing is greater than or equal to 50%, and otherwise as Normal.
- Statlog: This dataset has been 270 instance and 13 attributes. The missing attribute of dataset is null.

Table 1: Comparative result analysis of DT, NB, KNN and SVM using Accuracy, Sensitivity, Specificity and AUC with Hungarian dataset.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
DT[11]	89.2	89.4	89.6
NB[13]	89.1	89.2	88.9
KNN[25]	87.1	86.4	87.8
SVM[53]	79.6	78.2	80.7

Table 2: Comparative result analysis of DT, NB, KNN and SVM using Accuracy, Sensitivity, Specificity and AUC with Cleveland dataset.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
DT[11]	96.7	97.2	95.3
NB[13]	96.5	96.8	95.2
KNN[25]	89.6	87.7	90.9
SVM[53]	79.0	83.7	81.5

Table 3: Comparative result analysis of DT, NB, KNN and SVM using Accuracy, Sensitivity, Specificity and AUC with Z-Alizadeh Sani dataset.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
DT[11]	90.2	79.7	82.4
NB[13]	89.7	78.5	82.6
KNN[25]	88.6	78.2	83.3
SVM[53]	91.2	80.7	83.5

Table 4: Comparative result analysis of DT, NB, KNN and SVM using Accuracy, Sensitivity, Specificity and AUC with Statlog dataset.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
DT[11]	76.5	87.3	85.2
NB[13]	76.5	86.5	85.5
KNN[25]	79.2	87.7	80.4
SVM[53]	79.6	83.9	81.1

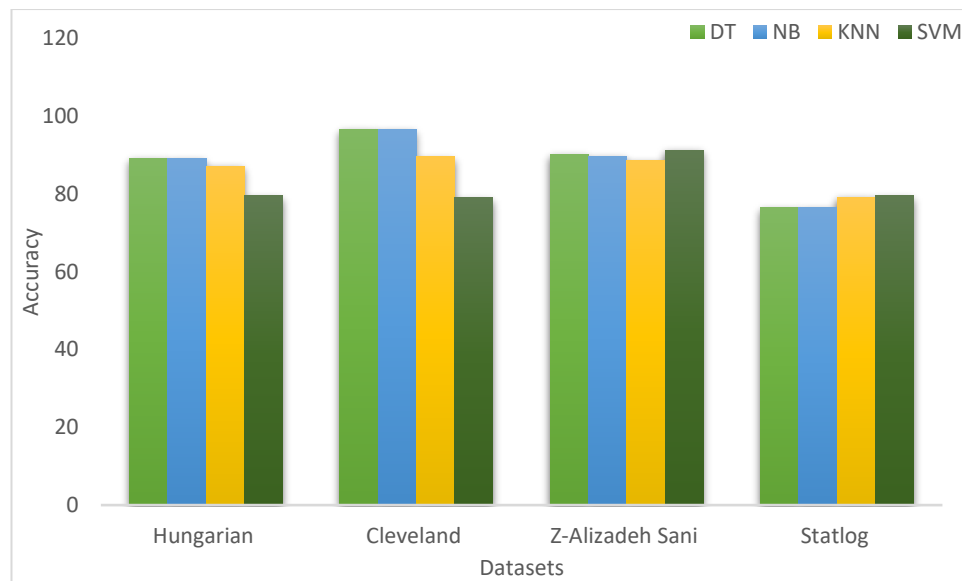


Figure 2: Comparative performance of DT[11], NB[13], KNN[25] and SVM[53] using Accuracy with Hungarian, Cleveland, Z-Alizadeh Sani and Statlog datasets. Here we observe the comparative better performance of accuracy: in case of Hungarian dataset for DT technique, in case of Cleveland dataset for DT technique, in case of Z-Alizadeh Sani dataset for SVM technique, in case of Statlog dataset for SVM technique,

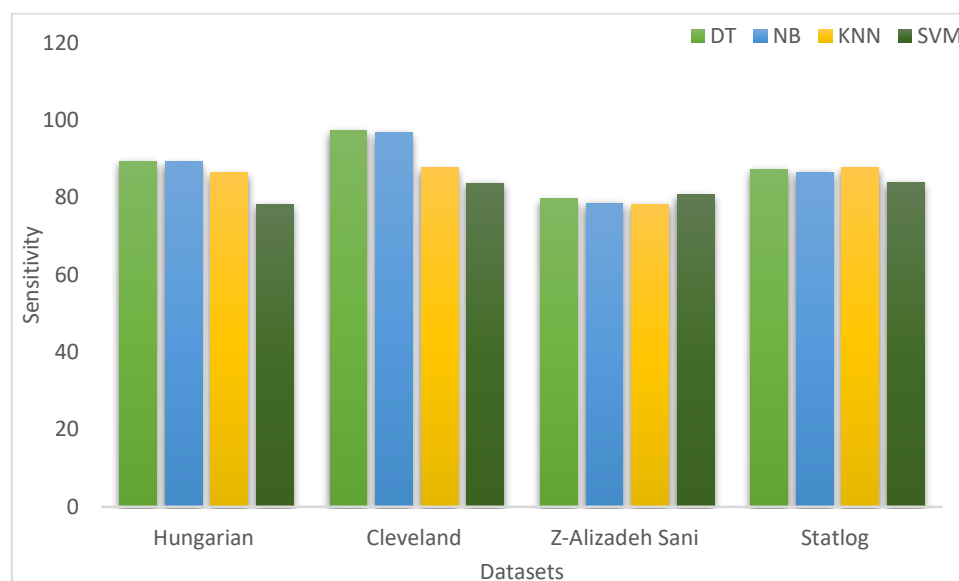


Figure 3: Comparative performance of DT[11], NB[13], KNN[25] and SVM[53] using Sensitivity with Hungarian, Cleveland, Z-Alizadeh Sani and Statlog datasets. Here we observe the comparative better performance of sensitivity: in case of Hungarian dataset for DT technique, in case of Cleveland dataset for DT technique, in case of Z-Alizadeh Sani dataset for SVM technique, in case of Statlog dataset for DT technique,

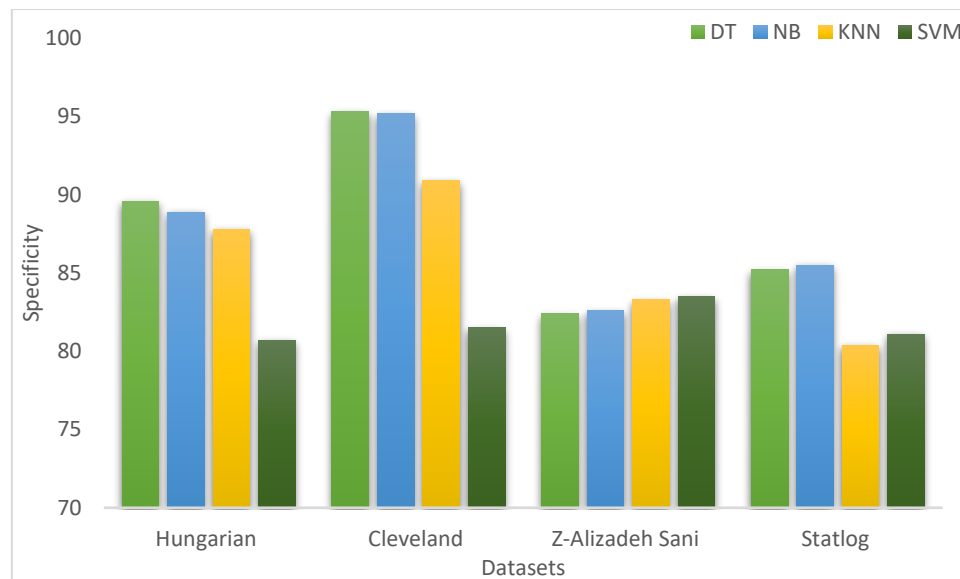


Figure 4: Comparative performance of DT[11], NB[13], KNN[25] and SVM[53] using Specificity with Hungarian, Cleveland, Z-Alizadeh Sani and Statlog datasets. Here we observe the comparative better performance of specificity: in case of Hungarian dataset for DT technique, in case of Cleveland dataset for DT technique, in case of Z-Alizadeh Sani dataset for SVM technique, in case of Statlog dataset for NB technique,

V. CONCLUSION & FUTURE SCOPE

In this paper analysed the performance of prediction of atherosclerosis diseases based on machine learning algorithm. The prediction and analysis of atherosclerosis disease machine learning applied four classification algorithm support vector machine, decision tree, naïve bayes and KNN. The results of support vector machine are influence the process of atherosclerosis disease detection. For the validation of algorithm use four data set such as Cleveland, Hungarian, Statlog and Z-Alizadeh Sani. The training and testing process applied 10 cross fold ratio. The NB and DT algorithms give better performance of KNN algorithms. The accuracy of KNN algorithms depends on the variation of K attributes. The selection of features attribute plays major role in prediction accuracy. The machine learning algorithms outcomes reached an overall accuracy of 88% and 85% of sensitivity. Compared with the support vector machine, decision tree, NB, and KNN where the accuracies respectively 87.00%, 89.01%, and 81.10%. Thus, this comparison has been shown that the support vector machine is better. Still accuracy of machine learning algorithms is challenging tasks. In future applied feature optimization algorithm for better selection of features and improved the accuracy of atherosclerosis prediction.

REFERENCE

- [1]. Terrada, Oumaima, Bouchaib Cherradi, Abdelhadi Raihani, and Omar Bouattane. "A novel medical diagnosis support system for predicting patients with atherosclerosis diseases." *Informatics in Medicine Unlocked* 21 (2020): 100483.
- [2]. Terrada, Oumaima, Bouchaib Cherradi, Abdelhadi Raihani, and Omar Bouattane. "Classification and Prediction of atherosclerosis diseases using machine learning algorithms." In *2019 5th International Conference on Optimization and Applications (ICOA)*, pp. 1-5. IEEE, 2019.
- [3]. Yilmaz, Nihat, Onur Inan, and Mustafa Serter Uzer. "A new data preparation method based on clustering algorithms for diagnosis systems of heart and diabetes diseases." *J Med Syst* 38 (2014): 48-59.
- [4]. Bhatla, Nidhi, and Kiran Jyoti. "A Novel Approach for heart disease diagnosis using Data Mining and Fuzzy logic." *International Journal of Computer Applications* 54, no. 17 (2012).
- [5]. Munger, Eric, John W. Hickey, Amit K. Dey, Mohsin Saleet Jafri, Jason M. Kinser, and Nehal N. Mehta. "Application of machine learning in understanding atherosclerosis: Emerging insights." *APL bioengineering* 5, no. 1 (2021): 011505.
- [6]. Parameswari, C., and S. Siva Ranjani. "Prediction of atherosclerosis pathology in retinal fundal images with machine learning approaches." *Journal of Ambient Intelligence and Humanized Computing* (2020): 1-11.

- [7]. Kolossváry, Márton, Carlo N. De Cecco, Gudrun Feuchtnet, and Pál Maurovich-Horvat. "Advanced atherosclerosis imaging by CT: radiomics, machine learning and deep learning." *Journal of cardiovascular computed tomography* 13, no. 5 (2019): 274-280.
- [8]. Singh, Navdeep, Punjab Firozpur, and Sonika Jindal. "Heart disease prediction system using hybrid technique of data mining algorithms." *International Journal of Advance Research, Ideas and Innovations in Technology* 4, no. 2 (2018): 982-987.
- [9]. Rao, V. Sree Hari, and M. Naresh Kumar. "Novel approaches for predicting risk factors of atherosclerosis." *IEEE journal of biomedical and health informatics* 17, no. 1 (2012): 183-189.
- [10]. Qawqzeh, Yousef K., Mohammad Mahmood Ootom, Fayez Al-Fayez, Ibrahim Almarashdeh, Mutasem Alsmadi, and Ghaith Jaradat. "A Proposed Decision Tree Classifier for Atherosclerosis Prediction and Classification." *IJCSNS* 19, no. 12 (2019): 197.
- [11]. Animesh Hazra, Subrata Kumar Mandal, Amit Gupta, Arkomita Mukherjee and Asmita Mukherjee. "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review." *Advances in Computational Sciences and Technology* 10, no. 7 (2017): 2137-2159.
- [12]. Shouman, Mai, Tim Turner, and Rob Stocker. "Integrating Naive Bayes and K-means clustering with different initial centroid selection methods in the diagnosis of heart disease patients." *CS & IT-CSCP* (2012): 125-137.
- [13]. Terrada, Oumaima, Bouchaib Cherradi, Abdelhadi Raihani, and Omar Bouattane. "Classification and Prediction of atherosclerosis diseases using machine learning algorithms." In *2019 5th International Conference on Optimization and Applications (ICOA)*, pp. 1-5. IEEE, 2019.
- [14]. Han, Donghee, Kranthi K. Kolli, Subhi J. Al'Aref, Lohendran Baskaran, Alexander R. van Rosendaal, Heidi Gransar, Daniele Andreini et al. "Machine learning framework to identify individuals at risk of rapid progression of coronary atherosclerosis: from the PARADIGM registry." *Journal of the American Heart Association* 9, no. 5 (2020): e013958.
- [15]. Nikan, Soodeh, Femida Gwady-Sridhar, and Michael Bauer. "Machine learning application to predict the risk of coronary artery atherosclerosis." In *2016 International conference on computational science and computational intelligence (CSCI)*, pp. 34-39. IEEE, 2016.
- [16]. Serrano, José Ignacio, M. Tomeckova, and Jana Zvárová. "Machine learning methods for knowledge discovery in medical data on atherosclerosis." *European Journal for Biomedical Informatics* 2, no. 1 (2006): 6-33.
- [17]. Georga, Eleni I., Nikolaos S. Tachos, Antonis I. Sakellarios, Vassiliki I. Kigka, Themis P. Exarchos, Gualtiero Pelosi, Oberdan Parodi, Lampros K. Michalis, and Dimitrios I. Fotiadis. "Artificial intelligence and data mining methods for cardiovascular risk prediction." In *Cardiovascular Computing—Methodologies and Clinical Applications*, pp. 279-301. Springer, Singapore, 2019.
- [18]. Kumar, Abhishek, Pardeep Kumar, Ashutosh Srivastava, VD Ambeth Kumar, K. Vengatesan, and Achintya Singhal. "Comparative Analysis of Data Mining Techniques to Predict Heart Disease for Diabetic Patients." In *International Conference on Advances in Computing and Data Sciences*, pp. 507-518. Springer, Singapore, 2020.
- [19]. Magesh, G., and P. Swarnalatha. "Optimal feature selection through a cluster-based DT learning (CDTL) in heart disease prediction." *Evolutionary Intelligence* (2020): 1-11.
- [20]. Rémy, Nfongourain Mougnotou, Tekinzang Tedondjio Martial, and Tayou Djamegni Clémentin. "The prediction of good physicians for prospective diagnosis using data mining." *Informatics in medicine unlocked* 12 (2018): 120-127.
- [21]. Repaka, Anjan Nikhil, Sai Deepak Ravikanti, and Ramya G. Franklin. "Design and implementing heart disease prediction using naive Bayesian." In *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 292-297. IEEE, 2019.
- [22]. Kazerouni, Faranak, Azadeh Bayani, Farkhondeh Asadi, Leyla Saeidi, Nasrin Parvizi, and Zahra Mansoori. "Type2 diabetes mellitus prediction using data mining algorithms based on the long-noncoding RNAs expression: a comparison of four data mining approaches." *BMC bioinformatics* 21, no. 1 (2020): 1-13.
- [23]. Shaji, Shaicy P. "Prediction and Diagnosis of Heart Disease Patients using Data Mining Technique." In *2019 international conference on communication and signal processing (ICCSP)*, pp. 0848-0852. IEEE, 2019.
- [24]. Atallah, Rahma, and Amjed Al-Mousa. "Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method." In *2019 2nd International Conference on new Trends in Computing Sciences (ICTCS)*, pp. 1-6. IEEE, 2019.
- [25]. Bashir, Saba, Zain Sikander Khan, Farhan Hassan Khan, Aitzaz Anjum, and Khurram Bashir. "Improving heart disease prediction using feature selection approaches." In *2019 16th international bhurban conference on applied sciences and technology (IBCAST)*, pp. 619-623. IEEE, 2019.

- [26]. Amin, Mohammad Shafenoor, Yin Kia Chiam, and Kasturi Dewi Varathan. "Identification of significant features and data mining techniques in predicting heart disease." *Telematics and Informatics* 36 (2019): 82-93.
- [27]. Dahiawade, Dhiraj, Gajanan Patle, and Ektaa Meshram. "Designing disease prediction model using machine learning approach." In *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 1211-1215. IEEE, 2019.
- [28]. Dhar, Sanchayita, Krishna Roy, Tanusree Dey, Pritha Datta, and Ankur Biswas. "A hybrid machine learning approach for prediction of heart diseases." In *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, pp. 1-6. IEEE, 2018.
- [29]. Dwivedi, Ashok Kumar. "Performance evaluation of different machine learning techniques for prediction of heart disease." *Neural Computing and Applications* 29, no. 10 (2018): 685-693.
- [30]. Rathnayakc, Bandarage Shehani Sanketha, and Gamage Upeksha Ganegoda. "Heart diseases prediction with data mining and neural network techniques." In *2018 3rd International Conference for Convergence in Technology (I2CT)*, pp. 1-6. IEEE, 2018.
- [31]. Goel, Sakshi, Abhinav Deep, Shilpa Srivastava, and Aprna Tripathi. "Comparative Analysis of various Techniques for Heart Disease Prediction." In *2019 4th International Conference on Information Systems and Computer Networks (ISCON)*, pp. 88-94. IEEE, 2019.
- [32]. Haq, Amin Ul, Jian Ping Li, Muhammad Hammad Memon, Shah Nazir, and Ruinan Sun. "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms." *Mobile Information Systems* 2018 (2018).
- [33]. Hasan, KM Zubair, Shourob Datta, Md Zahid Hasan, and Nusrat Zahan. "Automated prediction of heart disease patients using sparse discriminant analysis." In *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 1-6. IEEE, 2019.
- [34]. Hossain, Rifat, SM Hasan Mahmud, Md Altam Hossin, Sheak Rashed Haider Noori, and Hosney Jahan. "PRMT: Predicting Risk Factor of Obesity among Middle-Aged People Using Data Mining Techniques." *Procedia computer science* 132 (2018): 1068-1076.
- [35]. Maji, Srabanti, and Srishti Arora. "Decision tree algorithms for prediction of heart disease." In *Information and communication technology for competitive strategies*, pp. 447-454. Springer, Singapore, 2019.
- [36]. Maragatham, G., and Shobana Devi. "LSTM model for prediction of heart failure in big data." *Journal of medical systems* 43, no. 5 (2019): 1-13.
- [37]. Narayan, Subhashini, and E. Sathiyamoorthy. "A novel recommender system based on FFT with machine learning for predicting and identifying heart diseases." *Neural Computing and Applications* 31, no. 1 (2019): 93-102.
- [38]. Rady, El-Houssainy A., and Ayman S. Anwar. "Prediction of kidney disease stages using data mining algorithms." *Informatics in Medicine Unlocked* 15 (2019): 100178.
- [39]. Raju, Cincy, E. Philippsy, Siji Chacko, L. Padma Suresh, and S. Deepa Rajan. "A survey on predicting heart disease using data mining techniques." In *2018 conference on emerging devices and smart systems (ICEDSS)*, pp. 253-255. IEEE, 2018.
- [40]. Ramasamy, S., and K. Nirmala. "Disease prediction in data mining using association rule mining and keyword based clustering algorithms." *International Journal of Computers and Applications* 42, no. 1 (2020): 1-8.
- [41]. Raza, Khalid. "Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule." In *U-Healthcare Monitoring Systems*, pp. 179-196. Academic Press, 2019.
- [42]. Safdar, Saima, Saad Zafar, Nadeem Zafar, and Naurin Farooq Khan. "Machine learning based decision support systems (DSS) for heart disease diagnosis: a review." *Artificial Intelligence Review* 50, no. 4 (2018): 597-623.
- [43]. Sarkar, Bikash Kanti. "Hybrid model for prediction of heart disease." *Soft Computing* 24, no. 3 (2020): 1903-1925.
- [44]. Singh, Poornima, Sanjay Singh, and Gayatri S. Pandi-Jain. "Effective heart disease prediction system using data mining techniques." *International journal of nanomedicine* 13, no. T-NANO 2014 Abstracts (2018): 121.
- [45]. Woldemichael, Fikirte Girma, and Sumitra Menaria. "Prediction of diabetes using data mining techniques." In *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 414-418. IEEE, 2018.
- [46]. Palaniappan, Sellappan, and Rafiah Awang. "Intelligent heart disease prediction system using data mining techniques." In *2008 IEEE/ACS international conference on computer systems and applications*, pp. 108-115. IEEE, 2008.

- [47]. Yekkala, Indu, and Sunanda Dixit. "Prediction of heart disease using random forest and rough set based feature selection." *International Journal of Big Data and Analytics in Healthcare (IJBD AH)* 3, no. 1 (2018): 1-12.
- [48]. Hasan, S. M. M., M. A. Mamun, M. P. Uddin, and M. A. Hossain. "Comparative analysis of classification approaches for heart disease prediction." In *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*, pp. 1-4. IEEE, 2018.
- [49]. Kolukısa, Burak, Hilal Hacılar, Mustafa Kuş, Burcu Bakır-Güngör, Atilla Aral, and Vehbi Çağrı Güngör. "Diagnosis of Coronary Heart Disease via Classification Algorithms and a New Feature Selection Methodology." *International Journal of Data Mining Science* 1, no. 1 (2019): 8-15.
- [50]. Anitha, S., and N. Sridevi. "Heart disease prediction using data mining techniques." *Journal of Analysis and Computation* (2019).
- [51]. Diwakar, Manoj, Amrendra Tripathi, Kapil Joshi, Minakshi Memoria, and Prabhishek Singh. "Latest trends on heart disease prediction using machine learning and image fusion." *Materials Today: Proceedings* 37 (2021): 3213-3218.
- [52]. Malav, Amita, and Kalyani Kadam. "A hybrid approach for heart disease prediction using artificial neural network and K-means." *International Journal of Pure and Applied Mathematics* 118, no. 8 (2018): 103-10.
- [53]. Ismail, Ahmed, Samir Abdlerazek, and I. M. El-Henawy. "Big data analytics in heart diseases prediction." *Journal of Theoretical and Applied Information Technology* 98, no. 11 (2020).