International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6392
***Article Received**: 31 January 2023 **Revised**: 01 March 2023 **Accepted**: 07 March 2023*

CFLCA: High Performance based Heart disease Prediction System using Fuzzy Learning with Neural Networks

[1]Damodharan D, [2]Amit Kumar Goel, [3]Krishna Kant Agrawal, [4]Sheradha Johri, [5]Anuj Kumar
[1]Research scholar & Assistant Professor, Computer Science and Engineering
Galgotias University, Greater Noida, 201308, India
itechdamu@gmail.com
[2]Professor, Computer Science and Engineering
Galgotias University, Greater Noida, 201308, India
amit.goel@galgotiasuniversity.edu.in
[3] Professor,School of Computing Science and Engineering
Galgotias University, Greater Noida, 201308, India
krishna.agrawal@galgotiasuniversity.edu.in
[4] Research scholar,School of Computing Science and Engineering
Galgotias University, Greater Noida, 201308, India
Sheradha20@gmail.com
[5] Research scholar,School of Computing Science and Engineering
Galgotias University, Greater Noida, 201308, India
anujnsc@gmail.com

Abstract— Human Diseases are increasing rapidly in today's generation mainly due to the life style of people like poor diet, lack of exercises, drugs and alcohol consumption etc. But the most spreading disease that is commonly around 80% of people death direct and indirectly heart disease basis. In future (approximately after 10 years) maximum number of people may expire cause of heart diseases. Due to these reasons, many of researchers providing enormous remedy, data analysis in various proposed technologies for diagnosing heart diseases with plenty of medical data which is related to heart disease. In field of Medicine regularly receives very wide range of medical data in the form of text, image, audio, video, signal pockets, etc. This database contains raw dataset which consist of inconsistent and redundant data. The health care system is no doubt very rich in aspect of storing data but at the same time very poor in fetching knowledge. Data mining (DM) methods can help in extracting a valuable knowledge by applying DM terminologies like clustering, regression, segmentation, classification etc. After the collection of data when the dataset becomes larger and more complex than data mining algorithms and clustering algorithms (D-Tree, Neural Networks, K-means, etc.) are used. To get accuracy and precision values improved with proposed method of Cognitive Fuzzy Learning based Clustering Algorithm (CFLCA) method. CFLCA methodology creates advanced meta indexing for n-dimensional unstructured data. The heart disease dataset used after data enrichment and feature engineering with UCI machine learning algorithm, attain high level accurate and prediction rate. Through this proposed CFLCA algorithm is having high accuracy, precision and recall values of data analysis for heart diseases detection.

Keywords- Classification, Clustering, D-Tree, Data mining, Heart Disease, K-Medoids, K-means Algorithm, k- Nearest Neighbor, Naive Bayes,Neural Network, Regression.

I. INTRODUCTION

A. Knowledge Data Discovery (KDD)

The main reason to increase the capabilities of human analysis is to handle extensive number of bytes which are collected from both scientific and economic. All business utilizes the data mainly to obtain the advantage of competitions and increment the efforts to give useful services which can be helpful for the customers. Data which are captured from the surrounding are said to be the fundamental proof one uses for constructing various models as well as theories. Since PCs have allowed humans to easily access large amount of information that one can accumulate, it is normal to convert into computational techniques just to derive a valuable structures and patterns from large volume of data. Subsequently, KDD is mainly used to express an issue that is made by digital information which is an unacceptable fact for every individual: data overload [1] [3].

KDD extracting the useful patterns from the data, data mining is considered as the most commonly used application of particular algorithm. The difference between the data mining steps and KDD process is a main point which is being discussed

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6392
***Article Received**: 31 January 2023 **Revised**: 01 March 2023 **Accepted**: 07 March 2023*

in this paper. The steps that are involved additionally in the KDD process are data cleaning, data selection, data readiness, incorporating prior knowledge which is appropriate and legitimate mining results interpretations, which are necessary to make sure that helpful information/data are obtained. We must be careful while considering the applications of data-mining methods because applications taken blindly may be a very risky, effectively gives invalid patterns and meaningless information. KDD has developed and it is kept on developing day by day by combining all the research fields all together, such as the design recognition, databases, machine learning, statistics, knowledge procurement for master systems, AI, data visualization, and elite registering. The main objective is to extract the abnormal state of information from the low-level data from extensive data sets. The component of data mining in KDD depends totally on known techniques which include design recognition, machine learning, and statistics to discover useful patterns from the datasets in the process of KDD. KDD are not quite the same as example recognition or machine learning (and related fields). KDD mainly used for, how the data are accessed as well as stored, how in the massive datasets the algorithms are applied and then also it works proficiently, how we can interpret and envision the results, and how the interaction between the man-machine may be supported and modeled [2]. The KDD process contains multiple activities that use techniques for a specific discipline such as machine learning. In this paper, to contribute to KDD many opportunities have been found in the fields of different AI (except for machine learning). KDD gives an uncommon accentuation which finds the patterns in understandable form that are represented as helpful and intriguing information. Data warehousing is a related field that comes out from databases, for decision support and online analysis the well-known trends of businesses like collecting and cleaning transactional data is made accessible. The platform of KDD is set by the data warehousing in two ways such as data cleaning and the data access [4].

Data Cleaning: It is mainly used when the database contains all irrelevant, noise and errors, consistent data as well as missing data. To make this irrelevant data as relevant data, also it is performed to obtain useful information.

Data Access: It is capable enough to access the data from the database, and well-known methods should be made just for retrieving the data from the database and provides a path to access the data that are very hard to obtain (for example offline storing data).

KDD process includes various steps which includes the decision of many clients they are as follows:

1. In this step, it basically improves the understanding level of the applications for each and every individual, and also it determines the main goals of KDD process from the views of the customer.
2. In this step, the discovery is mainly performed on the targeted data sets in which we first select the required data set and then we concentrate more on the data samples.
3. It basically discusses about two things that is data cleaning and data preprocessing. It eliminates all the noise and missing data that may be found from the database through data cleaning. After that it performs data preprocessing in which it converts the raw data into easy an understandable form [5].
4. Data reduction and projection are mainly discussed. Depending on the main objectives of the task, it describes useful characteristics to represent data. It generally uses reduction method to reduce the attributes by decreasing the data in the database.

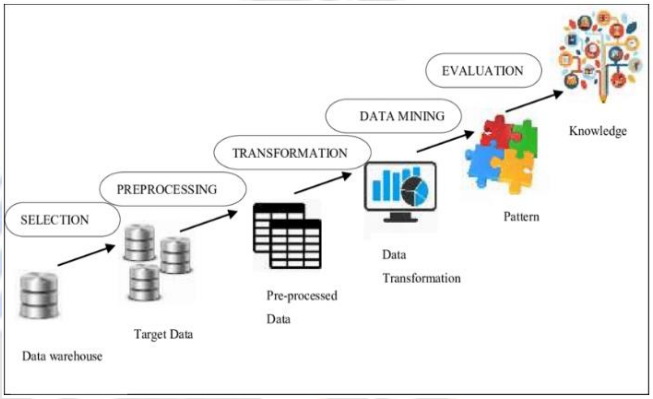

Figure 1: Knowledge Data Discovery Process

5. Deals with selecting the appropriate tasks of data mining. It is quite similar with the goals of (step 1), it recognizes the methods like clustering, classification and regression which helps in deciding whether these methods are the main goals of KDD process or not.
6. This step is responsible for choosing the algorithms of data mining. For searching a useful pattern in a data set, few methods would be selected. It also tells those which parameters and models usage would be more appropriate. It also compares the entire norms of KDD process with the specific data mining methods [6].
7. Searching of patterns (classification rules or trees, clustering and regression etc.)
8. Used for clarifying the patterns which are already mined. This step includes the models as well as the extracted patterns of visualization.
9. It mainly deals with enhancing the discovered knowledge. In this process, with the help of already accepted

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6392
***Article Received**: 31 January 2023 **Revised**: 01 March 2023 **Accepted**: 07 March 2023*

information it additionally incorporates inspecting and settling potential conflicts

Predictive task includes the methods within it such as classification, regression, time series analysis and prediction. The description focuses more on describing the data patterns which can be easily understood by human. The descriptive task includes clustering, summarization, association rules and sequence discovery.

1) Data Mining Methods

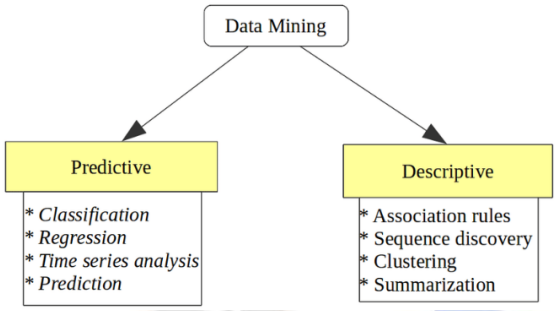

Figure 2: Data Mining Tasks

The boundaries in between the predictive and descriptive models are found not so strong. For a particular data mining application, the importance of the prediction and description may differ some times. The goals of description and prediction are accomplished by using variety of mining methods of data [7][8].

Classification: In classification, we generally predict the result of anything in the form of discrete outputs. In a simple way we can just say that it is nothing but mapping the input variables in the discrete form. The target class in the data for each case is easily predicted with full accuracy, which is referred as one of the important goals of classification. For example, in giving loans to the applicants the bank keeps the records of high, medium and low risk of credit by classifying all the aspect loan is given to the deserving applicant. So, in this way the classification is used everywhere to get a successful result.

Regression: Regression is the function of data mining which predicts the result based on continuous outputs. It is also used for the numerical predictions such as age, weight, distance and time etc. It mainly starts with the known target values of data set. For example, the heights of the children can be predicted numerically by observing data over the period of time for all the children. Regression always comes under the supervised learning.

Clustering: Clustering can be defined as grouping of similar data into a single group is said to be called as a cluster. Each clusters have a different group based on the similarities of the objects. Basically, clustering is done to reduce data size by reducing its complexity based on some attributes. For example, we are having thousands of essay writing with us and if we need to assemble them in some order then surely it would become a very complicated task for us. By the uses of clustering this problem can be solved easily by using some attributes like word frequency, page count and length of sentence, based on these attributes the essays can be made into different cluster forms and hence the problem is resolved quickly.

Summarization: Summarization is mainly used to represent the mean and standard deviation in the tabular form supports for analysis and mathematical functions about data. Clustering always summarizes the large sets of data. Summarization comes under the descriptive models in the methods of data mining.

Dependency modeling: Dependency modeling includes the model that determines remarkable dependencies between the variables. A dependency model occurs in two levels:

Level 1: The model of structural level represents the certain variables which are dependent on each other.
Level 2: The model of quantitative level which determines the dependencies of the strengths by utilizing some numeric measures.

Change and deviation detection: The most notable changes in the data that would be measured beforehand and it will be discovered by focusing more on it with the help of deviation detection and change [9].

B. **Cluster Analysis**

On the basis of the properties of a specific set of objects, the division of these objects on the basis of their similarities is known as clustering process. A specific joint algorithm that can be applicable to almost all required information analysis is done with the help of the technique which partitions the data. An object is made to be a part of the cluster or is assigned to it with the help of the clustering analysis method. The type of grouping involved here is the hard partitioning. Within the determination degree each object is related to a cluster. This is known as the soft partitioning method. There are varieties of divisions proposed within the clusters that are based on the distinct objects present. The partitioning method also depends on the various models present. On the basis of the relationship amongst the objects and the organization present there are various models proposed [11].

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6392
***Article Received**: 31 January 2023 **Revised**: 01 March 2023 **Accepted**: 07 March 2023*

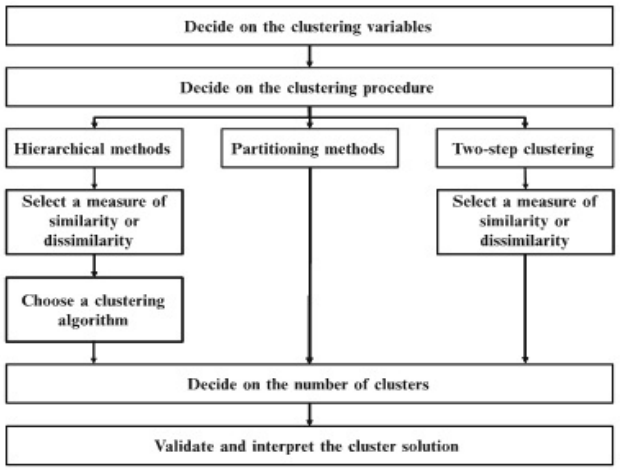

Figure 3: Steps in a Cluster Analysis

1)Clustering Analysis Techniques

Hierarchical: The representation of clusters in these methods is done at different levels of granularity, using dendrogram. The hierarchical system representation can be either assortment or isolate, on the basis of hybrid/hierarchical model. It can be either top-down or bottom-up designed.

Agglomerative: However, a sampling of data points amongst two clusters is utilized in sampled linkage. This is used to calculate the average distance of the two clusters. The centroid-linkage process utilizes the distance between the centroids [13]. In the terms of the depths of various nodes or a tree there is no special requirement of having a perfect balanced tree where the degree of each branch is exactly two. Various tradeoffs are provided for the balancing of node depths and node weights to construct a tree structure.

Density and Grid-Based Methods

DBSCAN and STING are the two classical techniques utilized for the density-based and grid-based techniques. Amongst the data points within a continuous space, the density-based methods are naturally defined. This is a very tough task for the density-based methods. Thus, within the discrete or non-Euclidean space, there cannot be a meaningful utilization. For this purpose, an embedded approach is to be utilized. Without any specialized transformations, it is tough to utilize the various arbitrary data types such as the time-series data within the density-based methods. In case of higher dimensionality, the density computations are extremely difficult to define due to the higher number of cells within the grid structure along with presence of sparse data in the grid available [14].

Flat: The data division is done and various clusters are formed with the help of certain partitioning representatives in this case. It is important to partition the representative and distance functions and regulate the behavior of the algorithm. Towards the nearest representatives the data points are assigned on each iteration. Iterative nature of EM algorithm is compared with this technique as there are soft tasks performed in each E-step and model parameters are adjusted within the M-step. There are various methods which help in creating the partitions which are described in the section below:

1. **k-Means:** The mean values of vector estimation of the partitioning representatives within these techniques. There is no original data set from which the partitioning representative is drawn. It is designed as the function of the data present. For the purpose of computing the distances, the Euclidean distance is utilized. One of the simplest techniques for the purpose of clustering data is the k-means method. Due to its simple nature this method is used most widely in the practical implementations.

2. **k-Medians:** For the purpose of creating the partitioning representative, the median within each dimension is utilized within these techniques instead of using the mean. From the original data set the partitioning representatives are not drawn in k-means technique. There is high sensitivity of the median of a set of values due to the extreme values present in the data in the case of k-medians approach. Hence, this technique is more stable to noise as well as outliers. The partitioning representatives are drawn from the original data in the case of k-Median technique which is otherwise also known as k-Medoids technique. However, both these techniques are not the same and have variations within them [15].

3. **k-Medoids:** From the original data present, these methods sample the partitioning representative. The cases which involve the clustering of data points which are arbitrary objects, these techniques are involved. The functions of these objects are not be much discussed here. For instance, discussing about the mean and median of a set of network or discrete sequential objects is not meaningful. In these situations, from within the data the partitioning representatives are achieved. The iterative methods help in enhancing the quality of those representatives. From within the representatives, one representative is replaced from within the current data of each iteration. This helps in determining whether the quality of clustering is enhanced or not. This method is thus considered to be as of the hill climbing method. As compared to the k-means and k-medoids techniques, these methods need more iteration. The situations in which the discussion of means or medians of data objects are not meaningful; this method can be utilized. This is however, not possible in the case of the other two methods.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6392
***Article Received**: 31 January 2023 **Revised**: 01 March 2023 **Accepted**: 07 March 2023*

C. ***Prediction with different DM techniques for Heart Diseases***

Data mining algorithms are utilized for extracting the hidden patterns in the enormous datasets to obtain valuable information. Vast amount of data is generated day by day by utilizing diverse symptoms of patients as well as the clinical reports in the medical fields. Therefore, data mining is utilized vigorously for obtaining valuable information from the huge database. For the clinical diagnosis, the explored hidden pattern from the database is utilized. Notwithstanding, now a days the medical datasets are widely increasing, which are heterogeneous in nature. With the help of the hospital management these datasets should be combined and sorted out.

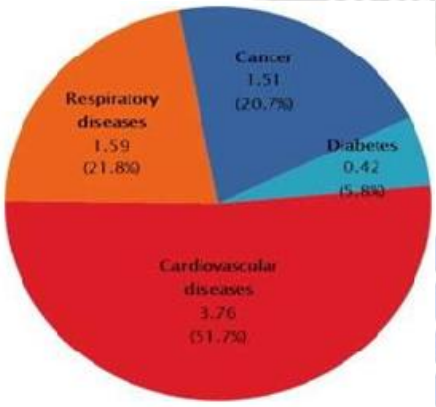

Figure 4: Phi Chart of Heart Diseases

Heart diseases are most common disease that is spreading rapidly in the today's generation. As per world health association around 12 million people died due to this heart problem. This disease is occurring among any age of people due to their poor diet, hereditary and life style which means that people's life is at risk. As it is increasing regularly so it needs to be analyzed properly and accurately. It can be analyzed utilizing the help of medical specialist. Combination of various techniques and medical system will provide high level predictive information. The risk factors can be described in the form of this chart:

From the past experience it has been noticed that algorithms and complex data are of great benefits with the existing package and software. This technique is commonly utilized in the various fields like crime analysis, engineering, medicines, prediction of expert, portable computing and mining in field of web. Medical diagnosis should be predicted accurately because it is very complicated as well as critical task. Data mining is always considered to be an important as well as necessary step of knowledge discovery. To explore the relationships and also the hidden patterns from the given database, the data mining had to join with the machine learning, database innovation and statistical analysis.

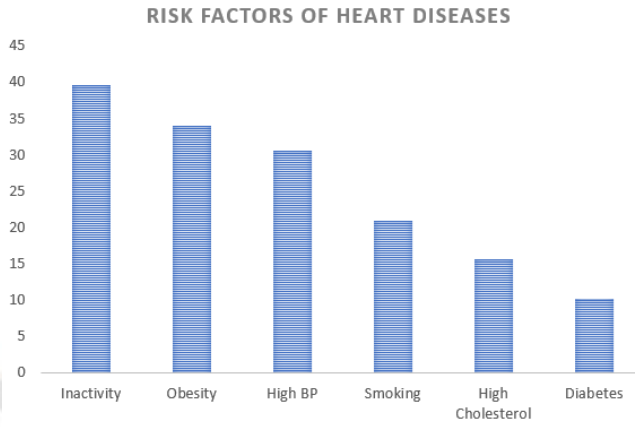

Figure 5: Risk Factors of Heart Disease

Data generally uses two kinds of learning such as unsupervised as well as supervised learning. Every technique of the data mining serves as an alternate based on the objectives of the modeling. Prediction and classification are the common objectives of modeling. [1] Prediction of categorical marks (discrete, unordered) is done in classification models while prediction of continuous-valued functions is done in prediction models. Naive Bayes classifier, networks of neural, decision tree and k mean algorithm, etc., giving various accuracy and performance analysis for any medical predication (heart infection treatment).

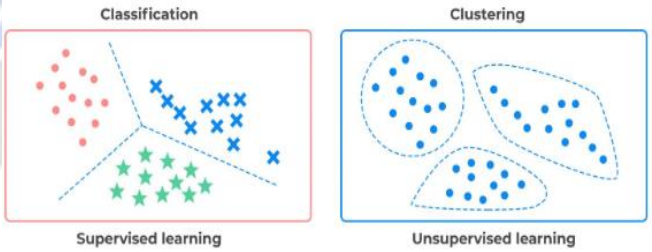

Figure 6: Supervised & Unsupervised Learning

D. Algorithms used in Heart Disease Prediction

Distinctive algorithms of supervised learning that is Neural Network, Naive Bayes, Apriori calculation of association, calculating the decision for the dataset have been utilized in this study. Weka 3.6.6 which is one of the data mining tools is utilized in testing. It is likewise appropriate for developing new machine learning schemes. List of few ML schemas as Neural networks, Naïve bayes, kNN, D-tree, etc.,

II. LITERATURE REVIEW

Ranganatha.S et.al "Analysis for Heart Disease and Medical Data Mining Dataset using Techniques of Classification, 2013": A modern medicine in today's world generates a lot of information that are stored in database. Data mining was used to solve the problem of heart disease. The management of hospital information can be improved very easily. Primarily the patient care activity was managed by the medical fields and it is also

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6392
***Article Received**: 31 January 2023 **Revised**: 01 March 2023 **Accepted**: 07 March 2023*

directed to the resources of the [12]. The main motive of gathering these data from the field of medicines is to get helpful for each and every patient. The study mainly stores the information of the patients who are suffering from heart disease and need to be hospitalized. The algorithms of data mining such as naive bayes and ID3 will be utilized when data sets are large or complex. ID3 outputs the outcome as decision tree which can be effectively caught on. The chances of heart diseases are predicted by utilizing the Bayesian for a given condition. Because of the immeasurable volume of information that is generated, people developed few algorithms that deliver output for a given query. The objective of this paper was to store large chunks of data and it gives Smooth workflow.

K.Srinivas et al, Diagnosis of Heart disease medically is very confusing task that ought to be accurately performed. The survey helps in analyzing the study in aspect of risk rates among the different age of people [15]. Subordinate several incorporate measures that are self- reported of being determined to have disease of cardiovascular such as (1) stroke (2) chest pain (3) heart attack. The morbidity is predicted by the study of heart indicating many attributes. The system that is automated enhances care provided by medical centers and reduces cost for medical diagnosis. In this research work, well known techniques of data mining specifically, making decision tree, using Naive Bayes and Networks of Neural are utilized in the treatment of heart disease. This can advance enhanced and further expanded. Fundamentally the list of attributes mentioned in database of medical center is utilized for predicting heart attack. Other than this rundown, we need to combine different types of attributes that has an impact on the output like person's financial status, pollution, tensions and past history of any medical record. Other than the techniques of data mining, clustering, series in aspect of time and rules of association are additionally utilized to judge the behavior of the patients. It is therefore used to study the condition of patient's morbidity in the health care centers.

R. Kavitha et.al, An advanced DM framework used in prediction with heart disease Classification using ANN based feature selection and feature extraction technique,2016: In order of high dimensional data set in the database of heart diseases is utilized as a part of the data mining stage of pre-processing. The crude database comprises of irrelevant and conflicting information subsequently incrementing the pursuit data storage data. One needs to remove the irrelevant and redundant data that are present, to achieve the characterization accuracy. The strategy of reduction is utilized to combine the data from higher to lower dimensional data with a few variables. For the simple heart disease prediction, a structure is integrated [16].

The structure made by utilizing the essential segments to extract the features and numerical model is processed to choose the pertinent characteristics utilizing the significant requirement. The work proposed, can be used to enhance the accuracy, effectiveness, and process having speed. The study may be used in image processing, recovery of data, and coordinating patterns. The subnet of component is selected and utilized by utilizing wrapped channel classifier to give appropriate output. When compared with other scoring function like Pearson correlation coefficients and Euclidean distance, improving the performance of the system. It facilitates later on efforts the exception will be addressed by demonstrating better level of accuracy.

Eman Abu Khousa et.al "Predictive Data Mining to Support Clinical Decisions: An Overview of Heart Disease Prediction Systems,2012": The Organization of health care is facing with many problems to give high quality of care to the patients in less cost. The volume of data that are accessible from the healthcare information systems in the database will be analyzed by both clinicians and administrators with a specific end goal to invent learning and settle down based on some decisions. It is mainly used to improve the diagnosing level of disease and its prevention is also improved [17]. It is the fate if there arises an occurrence of diseases related to heart which will be viewed as a major cause of death in individual. It utilizes analysis tools to discover the valuable patterns as well as hidden relationships in medical data. All systems display some strengths as well as drawbacks regarding the kind of data it handles, accuracy, interpretation, simplicity of, speculation ability and reliability. Poor speculation is still a noteworthy problem in healthcare basically as a result of shortage of information. In this research work, we actually studied regarding the systems that include five several techniques of data mining support to predict patients suffering level of heart diseases. Each effective models from those of five takes out a valuable pattern for the prediction such as drawbacks and strengths, interpretation of simplicity, data accuracy and speculation ability and reliability.

M.Gandhi et.al [10] Predictions in Heart Disease using Techniques of DM: A vast amount of information is delivered in medical affiliations but data is not appropriately utilized. The healthcare systems are fully fledged in aspects of data storing but however poor in matter of fetching data. To discover connections and patterns, the successful analysis methods are found to be absent in healthcare systems. Several methods are used in data mining which acts as a solution in this circumstance. This research work using for discover utilization parts for aid of people in healthcare systems by strategy for machine adapting moreover procedures of data mining [18]. The primary points basically recommend the system which is automated for diagnosing heart diseases by considering prior information and data. This paper gives information regarding the different knowledge abstraction techniques by utilizing methods of data mining which are being utilized for prediction of heart disease in today's research. In this research work, methods of data mining to be specific by using algorithm of decision tree, Naive Bayes and Networks of Neural will be analyzed on medical data sets.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6392
***Article Received**: 31 January 2023 **Revised**: 01 March 2023 **Accepted**: 07 March 2023*

Each and all data mining techniques have advantages as well as disadvantages of classification of data and extraction of knowledge. Moreover, decision tree, Networks of Neural or Naive Bayes can be researched in deep, which can be used in future in healthcare organizations.

Heon Gyu Lee et.al, The fundamental goal of this work is to build and after that intend another and methodology that are unique and are used in building the different characteristics of heart rate variability in treating the patients suffering from heart diseases. The people likewise intend appropriate prediction model to improve the treatments for cardiovascular disease [19]. For three recumbent postures the HVR is analyzed. The collaboration impacts in between the groups of normal people and the recumbent postures and patients suffering from heart diseases were seen in light of Heart Rate Variability lists. The arteries are measured and utilized by the estimations of blood vessel which are very thick enough as different characteristics.

Heart disease patients experienced scanning of vein utilizing ultrasound devised of high resolution as part of a past study. Keeping in mind the end goal to extract different features, six classification methods are tested. For classification, the proposed different features are employed permitting to pick from a large pool of classifier all around concentrated on classification methods. Researchers propose the likelihood that features of multi-parametric, contemplating entire conceivable HRV features as a demonstrative tool might use in treatment of cardiovascular disease. Finally, a few methods of the supervised learning are considered which includes the Bayesian classifiers which are extended, CMAR, SVM and MDA. From the output of experiments, the methods of classifications such as SVM and CPAR outperform.

Limia Abed et.al "Using Data Mining Technique to Diagnosis Heart Disease,2012": Diagnosing the medical problems is a guarantee to makes use of the techniques of data mining. An expert helps in the treatment, which generally involves one human expert which may even commit some mistake. Interestingly the knowledge is extracted from chunks of clinical data by data mining and delivers the models of prediction and also utilizes the task of classification to get the treatment. Several methods are presented in to deliver the classifiers in the field of classification [20].

Naive Bayes is one of them. In this research work, previously executed experiment was discussed with the help of Naive Bayes technique keeping in mind the end goal to manufacture models of prediction which helps in an artificial treatment for the people suffering from heart disease in view of set of parameters contained in dataset which were measured already for the individuals. At that point compare the results with different techniques as indicated by utilizing similar to UCI repository data. The outcome that is derived considered as good from practical work, while every one of the answers were right. So, the model is accurately achieved by the ratio (100%). The outcome was matched with different performance similar to the dataset. Furthermore, another problem arises in the field of treatment, where every heart disease contains own parameters which combine together in diagnosing, and diseases that are totally different from each other. Therefore, in this research work, very good outcomes were presented in the diagnose of heart utilizing the images of Single Proton Emission Computed Tomography (SPECT) images with classifiers of naive bayes are compared with the other classifiers.

S.palaniappan et.al [11] Healthcare organization collects bulk of raw data from the healthcare system to extract hidden patterns to obtain valuable information that may be helpful in making standard decisions. Inventing of invisible patterns as well as relationships sometimes gets unexploited. To solve this problem, data mining techniques helps a lot to rectify this problem [21]. A prototype has been developed in this research papers utilizing the techniques of data, in particular, algorithm of decision Trees, Bayesian and Network of Neural. It demonstrates the outcome for every technique in understanding the defined mining to derive the objectives. It basically answers the complex "imagine a scenario where" which traditional decision support systems queries can't do. Medical profiles of the patients are utilized from the database like gender, patient's age, circulatory strain, and the glucose level which detects probability of occurring heart diseases among the patients. It gives the sufficient knowledge such as pattern recognition, deriving relationships between the factors of medical data identified with diseases related to heart which need to be set up. The IHDPS is considered as user-friendly, based on web, expandable as well as reliable which can run on .NET platform. IHDPS is improved and expanded in future. For instance, it puts together the attributes of other medical data. Therefore, it incorporates other techniques of data mining like clustering, Series of time and Rules of Association. Instead of using simply categorical data, continuous data is utilized. Mining of unstructured data that are available in the database from large chunks of data, text mining is utilized. Another great challenge is to integrate text mining as well data mining together.

S. Sivagowry et.al "An Empirical Study on Applying Data Mining Techniques for The Analysis and Prediction of Heart Disease,2013": In this research paper, the main focus was done in order to get proper dimensional data with the help of an advanced DM algorithm. When talking about health care centers, they are best at providing information, but the main problem arises in extracting this information and by data mining it is extracted. The main reason behind about extracting this is due to shortage of effective analyzing kind of tools. For extracting the information, here the data mining was used to give a good information. As the problem of heart and its disease is one of the biggest problems that is of great concern as various

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6392
***Article Received**: 31 January 2023 **Revised**: 01 March 2023 **Accepted**: 07 March 2023*

deaths are also caused by this. So, for this research used various hybrid techniques and ways to tackle this problem. In this paper the mining techniques as used to detect the heart problems. Here the technique of mining used was of classification type. In this, it played an important role and was such beneficial than the methods like the clustering, Regression and also the association type. In classification, the decision tree is used and was of greater much beneficial than Neural type of network and the Naive Bayes in various cases. The conclusion was that by the data mining techniques, it was one of the best methods in order to deal with the heart problems with lesser number of variables.

III. PROPOSED CFLCA ALGORITHM

This research work mainly concentrating on high level precision and recall based analysis for heart diseases detection and prevention. In general, the prediction analysis follows three type portions segmentation/feature extraction, cluster and classifier. In the first method segmentation/feature extraction which helps with define and segregate information based on the needs [2]. Second method is clustering technique of k-mean clustering is applied on heart disease dataset its outcome is calculates arithmetic mean of the input and centroid point. From centroid value will find the knowledge based average and vector (Euclidian distance). Then next clusters create based on similar values.

The clustered data is classified by applying classification technique. The performance of data classification depends upon the accuracy of clustering. In this work, the difficult task is to obtain string relationship between the attributes of the dataset. Moreover, here introducing proposed CFLCA, Cognitive Intelligence based Clustering Algorithm. CFLCA method improve the relationship between entities, strong relationship, an effective meta indexing and data dimensionality. Because of these CFLCA improves result prediction rate and accuracy level significantly.

A. OBJECTIVES:

- To propose improvement in k-mean clustering algorithm to increase accuracy in clustering.
- To obtain the string relationship between the attributes of the dataset.
- To get accuracy and efficiency in result using improved k-mean algorithm.
- To reduce the execution time in result to save time of any individual.
- To compare results in terms of accuracy, time in both proposed and existing techniques.

B. *RESEARCH METHODOLOGY:*

The target of this research work is to model a prototype to categorize patients during emergency condition by generating smart predictor for the clopidogrelist. The predictor model is a phase where the system is applied for predicting the emergency condition of the patients by measuring certain clinical parameters. This predictor model is the combination of enriched feature selection and CFLCA are clustered and classified by controlling complicities and characteristics measure. This model pretends to reduce the execution time for examining the emergency patients' condition waiting in the queue. The design intention is to use an effectual CFLCA is not used to measure threat during clustering and classification process. The medical information is extracted from the available IFTS data sheet from emergency department that comprises of text, classification and some constant information. The outcomes of precision are based on kNN that provides nearest neighbor values for clustering in an effectual manner as that of CFLCA using the triage analysis. It is known that, the anticipated CFLCA model is more effectual for the evaluation of available data and the other unknown data.

Here, the characteristics attained from Gaussian prototype are also used as categorizers. Also, it comprises of VPA system for examining the efficiency based on the pre-requisite execution time and precision. Based on the advancements encountered in wireless communication, the genomic information, e-health records, public-health related information and behavioral information for generating voluminous information is considered to be healthy with concealed patterns of big data analytics are demoralized to extract the perceptions and proof for reducing average healthcare cost as it produces effectual outcomes based on smart device usage. Based on the analysis, it is observed that there are various types of heart disease which is classified based on the type of severity. The simulation is done in MATLAB environment for analyzing the characteristics of AI based classifier and SVM classifier with the available tool box. The ultimate target is to compute the performance of the model for predicting heart disease based on the accuracy and precision values.

The metrics are generated and analyzed for measuring the severity of heart disease using MATLAB tools. Analysis is concentrated on the usage of various schemes relies on the available MATLAB and evaluating the outcomes of other tools. Here, ten-fold CV is used for classifying the patient's class using the analyzed characteristics. This method is considered to be more obvious based on RF for pre-processing and results are evaluated based on these techniques. The goal is to analyze the target of heart disease prediction model for cardiovascular patients with the use of various ML approaches. Here, certain recommendations are attained from choice-based hierarchy to retrieve effectual results with the desirable precision values by computing the performance based on sensitiveness and specificity. The execution for analyzing threat with developed heart disease predictor model along with the use of healthcare records evaluation is based on time-series analysis. The goal is

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6392
***Article Received**: 31 January 2023 **Revised**: 01 March 2023 **Accepted**: 07 March 2023*

to resolve drawbacks encountered in heterogeneous dataset for evaluating the threats over the stages of heart disease, i.e., 3 to 7.

The objective is to improve heart disease prediction with MATLAB for executing CFLCA that works effectually than existing probability-based approaches. This work helps in addressing the issues connected with patient's life and uses them to reduce precision against the anticipated statistical scheme. Various existing approaches make use of computerized disease prediction model using CFLCA and Artificial Neural Networks. It is performed with MATLAB environment by providing the inputs from UCI Machine Learning Repositories. The predictor model has to examine the severity of heart failure and to predict it in its preliminary stage. In this research various classification approaches like ANN, NB, Decision Tree and CFLCA are used for predicting the heart failure. The above-mentioned approaches are used for predicting the heart failure where the objective is to predict heart disease earlier. The most dominant features have to be selected to perform further classification and to predict the stages of heart disease of individual patients. The anticipated classification prototypes with various approaches comprises of concealed prototypes and optimal feature selection for evaluating and classifying heart disease patients. With the attained outcomes, it can provide enhanced precision over the dataset utilized for classification of data from the dataset as the inputs are attained from feature selection approaches. The automated disease predictor system is merged with Fuzzy Logic and kNN known as CFLCA which is used for evaluating the threats over the heart disease prediction. For automating heart disease evaluation, various iterations are performed. The results demonstrate that the recommender model is extremely suited for predicting the higher and lower chances of cardiac risks for the heart disease patients.

With the evaluation and examination, the purpose of this work relies over the benefits of the healthcare domain where ML approaches are used for analyzing the possibility of the disease. The research gaps need to be bridged and to be resolved by other schemes where the intention is to design CFLCA model for predicting heart disease with the grouping of results to identify significant classes of Cardiovascular individuals related to the heart disease outcomes. The characteristics and the features related to the high-risk heart disease patients are analyzed with other chronic disease like Cardiovascular is extremely complex. It is due to the fact that heart disease is highly connected with various other complications and some other diseases. The clustered object that comes under an identical class is known as cluster. These objects are clustered as one part and other objects are considered under other objects. The clustering process is to produce identical class objects by summarizing the objects over the cluster. The cluster evaluation comprises of partitioning the dataset into clusters with resemblances against label allocation to generate various clusters. The clustering process is more flexible to change and assists in developing helpful features.

The integration of Fuzzy Logic and Neural Network is described as an efficient system that is skilled with the use of precise learning process that is attained from CF foundations. The Learning is done with the data while the variations are performed with a local means. This work integrates CF with FIS. The inference system is constructed over the CF framework and the learning processes are used to accelerate the FIS metrics that are utilized as bias function. CFLCA strategy is based on implementing neural fuzzy on the UCI machine learning repository.

C. *Cognitive Fuzzy and Learning based Clustering Algorithm (CFLCA)*

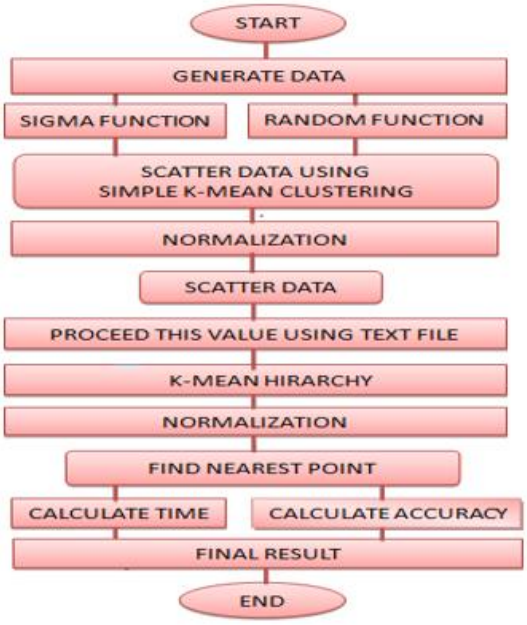

Figure 7: Flowchart of research methodology

1) Cognitive Fuzzy

This approach is well-known scheme where the Artificial Intelligence (AI) domain is utilized to predict appropriate information in every domain that is related to it. NFS function is based on training is tuned based on bias function by merging Fuzzy system. NFS is utilized to adjust bias function. It is alike of CFLCA and works like Fuzzy inference system based on its functionality. It is based on the Tsukamoto and Sugeno prototypes. This hybrid scheme is adjusted using CFLCA. It is specified with the fact that it comprises of two inputs $'x'$ and $'y'$ and the outcomes$'z'$. Based on this prototype model, the Sugeno Fuzzy model is produced with the following rules. General fuzzy model is designed by Takagi, Sugeno and Kang (1985). The rule format is provided as:

$$IF\ x\ is\ A\ and\ y\ is\ B\ THEN\ Z = f(x, y)$$

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6392
Article Received: 31 January 2023 Revised: 01 March 2023 Accepted: 07 March 2023

Here, AB are fuzzy sets in antecedents while $z = f(x, y)$ is a crisp function. FIS process under TS fuzzy method functions in the below given form:

Inputs fuzzification – Here, system inputs are made fuzzy

Using Fuzzy operator – fuzzy used to attain the output.

Rule format

Sugeno rule format is provided by an example below:

$$if\ 7 = x\ and\ 9 = y\ then\ output\ is\ z = ax + by + c$$

Condition 1: If $'x'$ specifies r_1 and $'y'$ specifies s_{p1} then $c_1 = a_1x + b_1y + k_1$

Condition 2: If $'x'$ specifies r_2 and $'y'$ specifies i_2 then $c_1 = a_2x + b_2y + k_2$

The working functionality of CFLCA framework is explained as figure10.

CFLCA network model comprises of two nodes: adaptable and fixed. These layers are composed of various other nodes predicted from node function. The anticipated model efficiency is based on adaptable parameters over nodes. The network learning functionality is based on certain parameter settings for reducing error over the appropriate output. It is depicted in Figure 10 (2 inputs and one output). The afore-mentioned rules are based on FIS method.

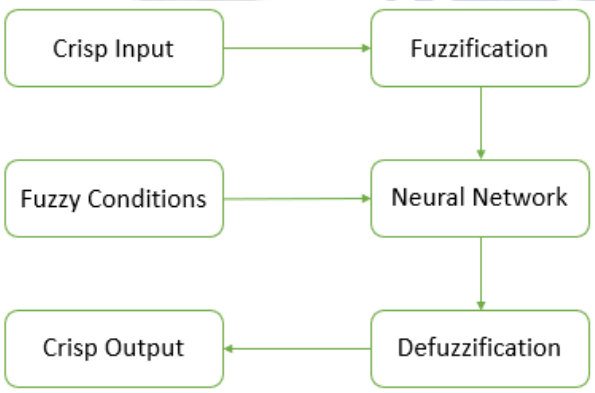

Fig 10: Hybrid neuro fuzzy

2) Cognitive Fuzzy and its learning methods

CFLCA ALGORITHM

Input: HeartDiseasesDatasetKM, dim n, CF=0, distance d, scalar_factor s

Output: SmartPredictionRate, ShieldDecisionRank

1. Procedure heart_disease_set_values (n, CF, d, s)
2. Read dataset and dimensionality of data collection has number of rows “y” and number of columns “z” total individual data_cell count “n”.
3. Construct heterogeneous entity based adjacent matrix
4. $HD = \begin{bmatrix} d_{exp}.CF_{act} & (1 - CF_{dact}).d_{dim,s} \\ (1 - CF_{act}).d_{dim,s} & CF_{act,s,dim} \end{bmatrix}$
5. loop(i=0; i=r-4; i++) /// selection of medoid point 1.
6. loop(j=0; j=m; j++)
7. Select s=data (i, j); End
8. Calculation of Euclidian distance()
9. loop(i=0; i=r;i++)
10. for initialized j compare with n values and in single increment
11. $i \rightarrow y_i^{(2)} = \prod_{i=1}^{k} x_{ji}^{(2)}$;
12. $j \rightarrow \sum_{i=1}^{n} \bar{\mu}_i[k_{io} + k_{i1} + k_{i2}]$;
13. $distance = \sqrt{(A(i+1) - A.i^2) - (B(j+1) - B.j^2)}$ End
14. Normalization () or backtracking ()
15. CFLCA(Input, kNN(Vec), Distance, Categories Threshold, decision)
16. For (k=0; k=data;k++)
17. Swap k(i+1) and k(i); else
18. Use k-1times with Repeat
19. Define(train, test, CFLCA)
20. Determine(precision, recall) iterate (k-1) end
21. repeat step 5 to 15 until all points get clustered.

a) Input Layer

This layer s termed as fuzzification layer where every node is considered as an adaptive membership function. The parameterized functions are membership function for fuzzy set or linguistic label which is of trapezoidal generalized bell or gaussian/triangular form, i.e., Gaussian membership function is depicted based on parameters couples (c, σ).

$$y_i^{(1)} = Gaussian\ (x: c, \sigma) = e^{\left(-\frac{1}{2}\right)\left(\frac{x-e}{\sigma}\right)^2} \quad (1)$$

Here, gaussian membership parameters are regularized by c, σ, these parameters are pointed as antecedent parameters, $y_i^{(1)}$ is output layer.

b) Hidden layer 1

This layer is known as product layer, where every node can react tosugeno fuzzy rules. Nodes are gathered from input with respective to neuron fuzzification. As an outcome, the firing strength of every rule is presented. The neuron output obtained from this latey is specified as in Eq. (2):

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6392
***Article Received**: 31 January 2023 **Revised**: 01 March 2023 **Accepted**: 07 March 2023*

$$y_i^{(2)} = \prod_{i=1}^{k} x_{ji}^{(2)} \qquad (2)$$

Here, $x_{ji}^{(2)}$ is input layer from $1(j)$ to $2(i)$ layer and output is specified as $y_i^{(2)}$ for neuron $'i'$ in product layer.

c) Hidden Layer 2

It is a standardized layer. It accepts feedback from product layer (neurons) and evaluates the weighted firing power. The neuron results are expressed as in Eq. (3):

$$y_i^{(3)} = \frac{x_{ji}^{(3)}}{\sum_{j=1}^{n} x_{ij}^{(3)}} = \bar{\mu}_i \qquad (3)$$

Here, $x_{ij}^{(3)}$ is received input and neurons generated from product layer to neuron over normalization layer $y_i^{(3)}$ is layer 3 output.

d) Hidden Layer 3

It is a defuzzification layer. The nodes are considered as adaptable or modifiable nodes. Neuron over defuzzification layer computes corresponding weights and values of certain rules are specified as in Eq. .4):

$$y_i^{(4)} = x_i^{(4)}[k_{i0} + k_{i1} + k_{i2}] = \bar{\mu}_i\,[k_{i0} + k_{i1} + k_{i2}] \qquad (4)$$

Here, $x_j^{(4)}$ is input of layer 4 while output is $y_i^{(4)}$. $k_{i0}, k_{i1}, k_{i2}$are successive parameters of $'i'$ rule.

e) Output Layer

This layer gives complete output for CFLCA model that integrates the output of previous layers. It is expressed as in Eq. (5):

$$y = \sum_{i=1}^{n} x_i^{(5)} = \sum_{i=1}^{n} \bar{\mu}_i[k_{io} + k_{i1} + k_{i2}] \qquad (5)$$

CFLCA learning process comprises of various updated parameters using two-pass learning, backward/forward pass algorithm. CFLCA parameters are trained for reducing the error among the desired and actual output.

EXPERMENTAL RESULT

In the first phase of clustering technique of k-mean clustering is applied in which heart disease dataset is given as input and calculated mean and standard deviation of segmented dimensional data. Then calculated centroid (center point) of each cluster with Euclidean distance and vectors also derived from the same. The similar data points are clustered in a single cluster and other are in the second cluster.

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
52	1	0	125	212	0	1	168	0	1	2	2	3	0
53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
61	1	0	148	203	0	1	161	0	0	2	1	3	0
62	0	0	138	294	1	1	106	0	1.9	1	3	2	0
58	0	0	100	248	0	0	122	0	1	1	0	2	1
58	1	0	114	318	0	2	140	0	4.4	0	3	1	0
55	1	0	160	289	0	0	145	1	0.8	1	1	3	0
46	1	0	120	249	0	0	144	0	0.8	2	0	3	0
54	1	0	122	286	0	0	116	1	3.2	1	2	2	0
71	0	0	112	149	0	1	125	0	1.6	1	0	2	1
43	0	0	132	341	1	0	136	1	3	1	0	3	0
34	0	1	118	210	0	1	192	0	0.7	2	0	2	1
51	1	0	140	298	0	1	122	1	4.2	1	3	3	0
52	1	0	128	204	1	1	156	1	1	1	0	0	0
34	0	1	118	210	0	1	192	0	0.7	2	0	2	1
51	0	2	140	308	0	0	142	0	1.5	2	1	2	1
54	1	0	124	266	0	0	109	1	2.2	1	1	3	0
50	0	1	120	244	0	1	162	0	1.1	2	0	2	1
58	1	2	140	211	1	0	165	0	0	2	0	2	1
60	1	2	140	185	0	0	155	0	3	1	0	2	0

Figure 10: K-means clustering

In figure 10, used advanced k-means clustering for find the nearest values and strong entity based dimensional of heart disease dataset. After preprocessing completed all data cleaning, data integration, data transformation and data deduction (feature engineering) process and got exact input for CFLCA Euclidian distance.

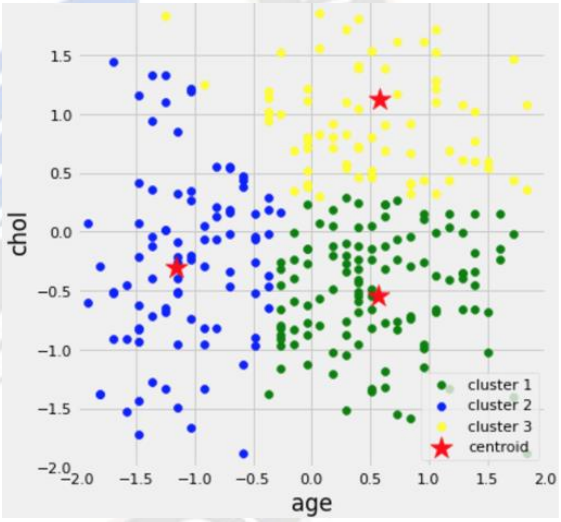

Figure 11: Clusters marked with centroid

In figure, CFLCA supports to improve the quality of the clusters with k-means algorithm and its normalization technique. An advanced clustering method improves the cluster quality using normalization and backtracking methods. Heart disease dataset for better analysis and prediction plotted on two-dimensional plane through CFLCA. Calculated center point of each cluster (centroid).

Final output of the CFLCA clustering algorithm:

From the above figure, to cluster the dataset k-mean clustering algorithm is used. In the algorithm central point is selected and Euclidian distance is calculated to generate final output. In the final output clusters are formed in which central points are highlighted

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6392
***Article Received**: 31 January 2023 **Revised**: 01 March 2023 **Accepted**: 07 March 2023*

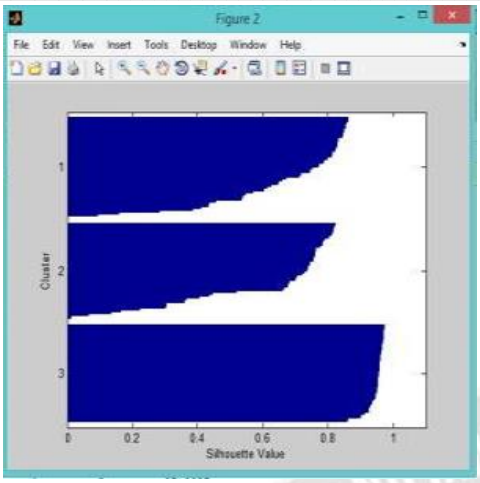

Figure 12: Cluster quality analysis

In figure, cognitive fuzzy learning based advanced clustering algorithm is used to segregate information with different dimensional and entities. Centroid (central point) is selected in the algorithm and Euclidian distance is calculated to generate final output. To analysis the cluster quality, graphs are drawn and it is analyzed that cluster quality of third cluster is low.

```
       id              age              sex              cp        
 Min.   :  1.0   Min.   :29.00   Min.   :0.0000   Min.   :1.000  
 1st Qu.: 76.5   1st Qu.:48.00   1st Qu.:0.0000   1st Qu.:3.000  
 Median :152.0   Median :56.00   Median :1.0000   Median :3.000  
 Mean   :152.0   Mean   :54.44   Mean   :0.6799   Mean   :3.158  
 3rd Qu.:227.5   3rd Qu.:61.00   3rd Qu.:1.0000   3rd Qu.:4.000  
 Max.   :303.0   Max.   :77.00   Max.   :1.0000   Max.   :4.000  
    trestbps          chol            fbs            restecg      
 Min.   : 94.0   Min.   :126.0   Min.   :0.0000   Min.   :0.0000  
 1st Qu.:120.0   1st Qu.:211.0   1st Qu.:0.0000   1st Qu.:0.0000  
 Median :130.0   Median :241.0   Median :0.0000   Median :1.0000  
 Mean   :131.7   Mean   :246.7   Mean   :0.1485   Mean   :0.9901  
 3rd Qu.:140.0   3rd Qu.:275.0   3rd Qu.:0.0000   3rd Qu.:2.0000  
 Max.   :200.0   Max.   :564.0   Max.   :1.0000   Max.   :2.0000  
    thalach          exang           oldpeak          slope       
 Min.   : 71.0   Min.   :0.0000   Min.   :0.00   Min.   :1.000  
 1st Qu.:133.5   1st Qu.:0.0000   1st Qu.:0.00   1st Qu.:1.000  
 Median :153.0   Median :0.0000   Median :0.80   Median :2.000  
 Mean   :149.6   Mean   :0.3267   Mean   :1.04   Mean   :1.601  
 3rd Qu.:166.0   3rd Qu.:1.0000   3rd Qu.:1.60   3rd Qu.:2.000  
 Max.   :202.0   Max.   :1.0000   Max.   :6.20   Max.   :3.000
```

Figure 14: Dataset loaded in K-mean Clustering:

In figure 14, the CFLCA framework improved clustering values and quality with normalization. This heart disease dataset is shows values with id, age, sex, cp, etc (16 entities).

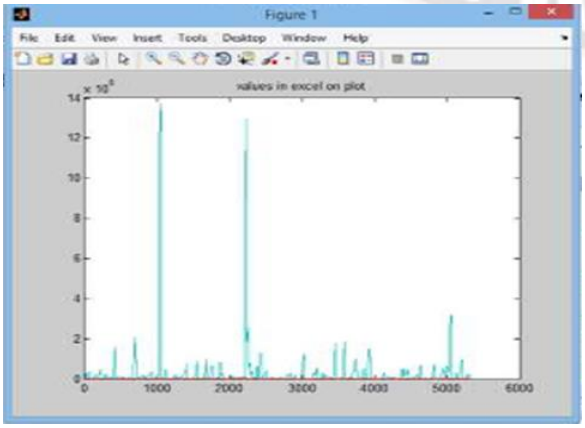

Figure 15: Plotted Dataset

Figure 15 shows that an advanced k-mean clustering is used to improves the data segmentation and classification. Because of prediction with experience (fuzzy learning). Figure 15 loaded and plotted two-dimensional representation of prediction rate and analysis with machine trained values.

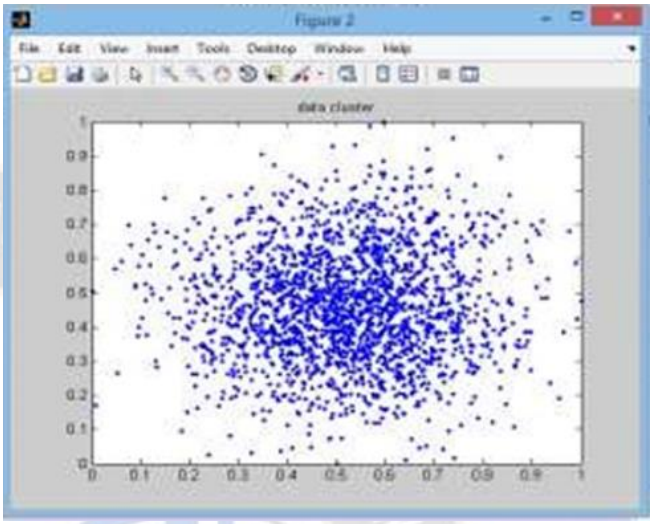

Figure 16: Data in Single Cluster

In figure 16, general analysis of heart disease dataset combined and plotted single two-dimensional plane itself as single cluster.

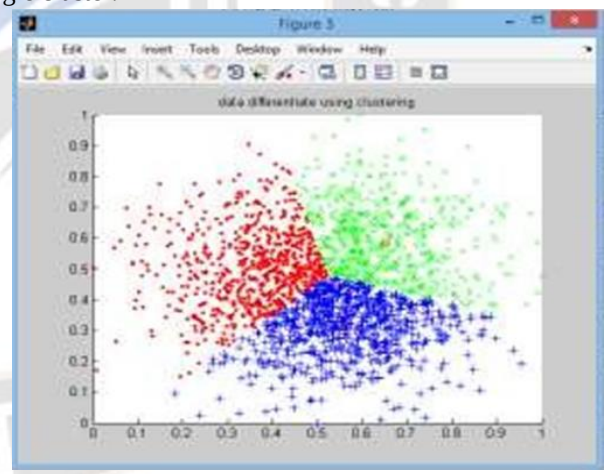

Figure 17: Data Differentiated Using Clustering

In figure explain variation-based entities and its dimensional processed with possible number of positive neutrons and plotter graph with three different clustering segments.
This plotted graph values represents all entity and factors are completely clustered using CFLCA algorithm. The central points (centroid) are marked in every cluster.

The parameter values which are obtained due to existing techniques are mentioned below: Accuracy = 93.24864 Time = 7.645456 GINI Index = 0.85681

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6392
***Article Received**: 31 January 2023 **Revised**: 01 March 2023 **Accepted**: 07 March 2023*

The Second Phase is Classification. The clustered data is classified by applying classification technique. The performance of data classification depends upon the accuracy of Clustering.

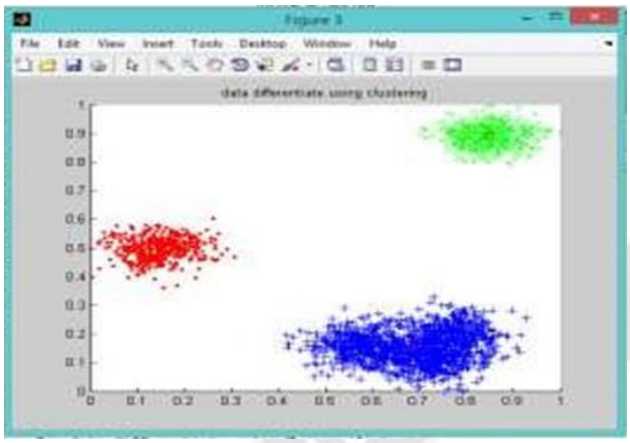

Figure 18: Data Clustered using K-mean clustering

COMPARISON WITH EXISTING TECHNIQUE

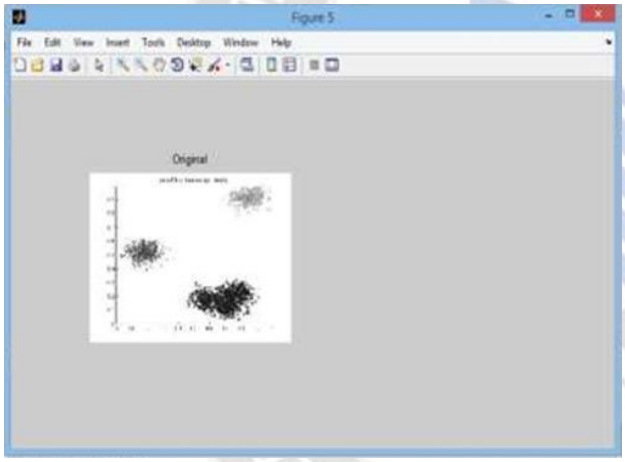

Figure 20: Accuracy & Execute Time of Base paper

Table 1 Precision Estimation for CFLCA

Precision	95%
Sensitivity	96%
Specificity	95.8%

Table 2 Comparison of performance measures

Metrics	**Hybrid DM algorithms**	**CFLCA**
Precision	88.7%	95%
Sensitivity	84%	96%
Specificity	82%	95.8%

The rate of classification error is 0.015. The precision of CFLCA model is depicted as in Table 1. Here, metrics like precision, sensitivity, and specificity is evaluated. The precision and sensitivity is 95%; while specificity is 95.8%. The anticipated model gives better results in heart disease prediction in earlier stage and assists the clopidogrel list to give proper treatment during the time of critical issues. Table 2 depicts the comparison of performance metrics like precision, sensitivity, specificity of hybrid DM algorithms and CFLCA respectively.

IV. CONCLUSION

This research work is used to detect heart diseases by using an advanced and optimal DM techniques by performing an effective and maximum rate of prediction analysis. These techniques help in digging meaningful information from large amount of data. In this paper K-mean clustering technique is utilized for the given datasets to calculate the athematic mean of input dataset from the center point. As the heart disease dataset becomes more complex nowadays, k mean algorithm can help in deriving a relationship between various entities of the dataset to predict the heart diseases accurately in health care organizations. This dimensionality of the attributes and its dataset can be reduced based on various machine learning algorithms with better time complexity in future.

Accuracy level can be increased by using some of the algorithms of data mining. Execution time is also reduced which helps the individual to save their time.

REFERENCES

[1] Fuyuan Cao, Joshua Zhexue Huang, Jiye Liang, Xingwang Zhao, Yinfeng Meng, Kai Feng, Yuhua Qian, "An Algorithm for Clustering Categorical Data With Set-Valued Features", IEEE Transactions on Neural Networks and Learning Systems, vol 29, issue 10, pp 4593 – 4606, 2018.

[2] Amjad Ali, Faiz Jillani, Rabbiah Zaheer, Ahmad Karim, Yasser Obaid Alharbi, Mohammad Alsaffar, Khalid Alhamazani, "Practically Implementation of Information Loss: Sensitivity, Risk by Different Feature Selection Techniques", IEEE Access, vol 10, pp 27643 – 27654, 2022.

[3] Hanan E. Abdelkader, Ahmed G. Gad, Amr A. Abohany, Shaymaa E. Sorour, "An Efficient Data Mining Technique for Assessing Satisfaction Level With Online Learning for Higher Education Students During the COVID-19", IEEE Access, vol 10, pp 6286 – 6303, 2022.

[4] Anjan Goswami, Ruoming Jin, and Gagan Agrawal, "Fast And Exact K-Means Clustering", International Conference on Data Mining (ICDM), 2004.

[5] Juhee Bae, Tove Helldin, Maria Riveiro, Sławomir Nowaczyk, Mohamed-Rafik Bouguelia, Göran Falkman, "Interactive Clustering: A Comprehensive Review", ACM Computing Surveys (CSUR), vol 53, issue 1, pp 1 – 39, 2021.

[6] Sangdi Lin;Bahareh Azarnoush;George C. Runger, "CRAFTER: A Tree-Ensemble Clustering Algorithm for Static Datasets with Mixed Attributes and High Dimensionality", IEEE Transactions on Knowledge and Data Engineering, vol 30, issue 9, pp 1686 – 1696, 2018.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6392
***Article Received**: 31 January 2023 **Revised**: 01 March 2023 **Accepted**: 07 March 2023*

[7] Fotis Savva, Christos Anagnostopoulos, Peter Triantafillou, Kostas Kolomvatsos, "Large-scale Data Exploration Using Explanatory Regression Functions", ACM Transactions on Knowledge Discovery from Data (TKDD), vol 14, issue 6, pp 1 – 33, 2020.

[8] Gowtham Atluri, Anuj Karpatne, Vipin Kumar, "Spatio-Temporal Data Mining: A Survey of Problems and Methods", ACM Computing Surveys (CSUR), vol 51, issue 4, pp 1 – 41, 2019

[9] Danyang Cao, Bingru Yang, "An improved k-medoids clustering algorithm", Computer and Automation Engineering (ICCAE), 2010

[10] Chen, H.L., Chuang, K.T and Chen, M.S., "Labeling Un clustered Categorical Data into Clusters Based on the Important Attribute Values", IEEE International Conference on Data Mining (ICDM), 2005.

[11] J.Zhu, H.Wang, M.Zhu, B.K.Tsou, and M.Ma,," Aspect-Based Opinion Polling from Customer Reviews," T. Affective Computing2(1):pp. 37- 49, 2011.

[12] M.Karamibekr,A.A.Ghorbani,"Verb Oriented Sentiment Classification," Processed of the IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (WI-IAT), Vol (1): pp.327-331, 2012.

[13] Yun Chi, Haixun Wang, Philip S. Yu, Richard R. Muntz, "Moment: Maintaining Closed Frequent Itemsets over a Stream Sliding Window," 2004, icdm, pp.59-66, Fourth IEEE International Conference on Data Mining (ICDM'04)

[14] Koh, J.-L., and Shieh, S.-F," An efficient approach for maintaining association rules based on adjusting FP-tree structures", 2004, ACm, research proceedings, pp. 53-123

[15] Ranganatha S., Pooja Raj H.R., Anusha C., Vinay S.K.," MEDICAL DATA MINING AND ANALYSIS FOR HEART DISEASE DATASET USING CLASSIFICATION TECHNIQUES", 2013, IEEE, 32987u-45096-45

[16] Syed Umar Amin, Kavita Agarwal, Dr. Rizwan Beg," Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors", 2013 IEEE Conference on Information and Communication Technologies (ICT) 978-1-4673-5758

[17] Theresa Princy, R, J. Thomas," Human Heart Disease Prediction System using Data Mining Techniques", 2016 International Conference on Circuit, Power and Computing Technologies [ICCPCT], 978-1-5090-1277-0

[18] K.Srinivas, Dr.G.Raghavendra Rao, Dr. A.Govardhan," Analysis of Coronary Heart Disease and Prediction of Heart Attack in Coal Mining Regions Using Data Mining Techniques", 2010, IEEE, 978-1-4244-6005-2

[19] R.Kavitha, E.Kannan," An Efficient Framework for Heart Disease Classification using Feature Extraction and Feature Selection Technique in Data Mining", 2016, IEEE, 29753-48506-65-7

[20] Eman AbuKhousa, Piers Campbell," Predictive Data Mining to Support Clinical Decisions: An Overview of Heart Disease Prediction Systems", 2012 International Conference on Innovations in Information Technology (IIT), 978-1-4673-1101-4

[21] Monika Gandhi, Dr. Shailendra Narayan Singh," Predictions in Heart Disease Using Techniques of Data Mining", 2015 1st International Conference on Futuristic trend in Computational Analysis and Knowledge Management (ABLAZE)

[22] Heon Gyu Lee, Ki Yong Noh, and Keun Ho Ryu," A Data Mining Approach for Coronary Heart Disease Prediction using HRV Features and Carotid Arterial Wall Thickness", 2008 International Conference on BioMedical Engineering and Informatics, 978- 0-7695-3118-2

[23] Lamia AbedNoor Muhammed," Using Data Mining technique to diagnosis heart disease", 2012, IEEE, 3545-56445-4

[24] Sellappan Palaniappan, Rafiah Awang," Intelligent Heart Disease Prediction System Using Data Mining Techniques", 2008, IEEE, 978-1-4244-1968-5

[25] Mai Shouman, Tim Turner, Rob Stocker," USING DATA MINING TECHNIQUES IN HEART DISEASE DIAGNOSIS AND TREATMENT", 2012, IEEE, 978-1-4673-0484-9

[26] Sivagowry .S, Dr. Durairaj. M and Persia.A," An Empirical Study on applying Data Mining Techniques for the Analysis and Prediction of Heart Disease", 2013, IEEE, 23543- 56-465

[27] K. Prasanna Lakshmi, Dr. C.R.K.Reddy," Fast Rule-Based Heart Disease Prediction using Associative Classification Mining", 2015, IEEE International Conference on Computer, Communication and Control (IC)

[28] Elena Baralis, Silvia Chiusano, and Paolo Garza," A Lazy Approach to Associative Classification", 2008, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, Vol. 20, No. 2

[29] K. Prasanna Lakshmi, Dr. C. R. K. Reddy," Compact Tree for Associative Classification of Data Stream Mining", 2012, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2.

[30] Bing Liu, Wynne Hsu, Yiming Ma," Integrating Classification and Association Rule Mining", 2012, Research gate, AAAI organization, pp. 345-753.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6392
***Article Received**: 31 January 2023 **Revised**: 01 March 2023 **Accepted**: 07 March 2023*

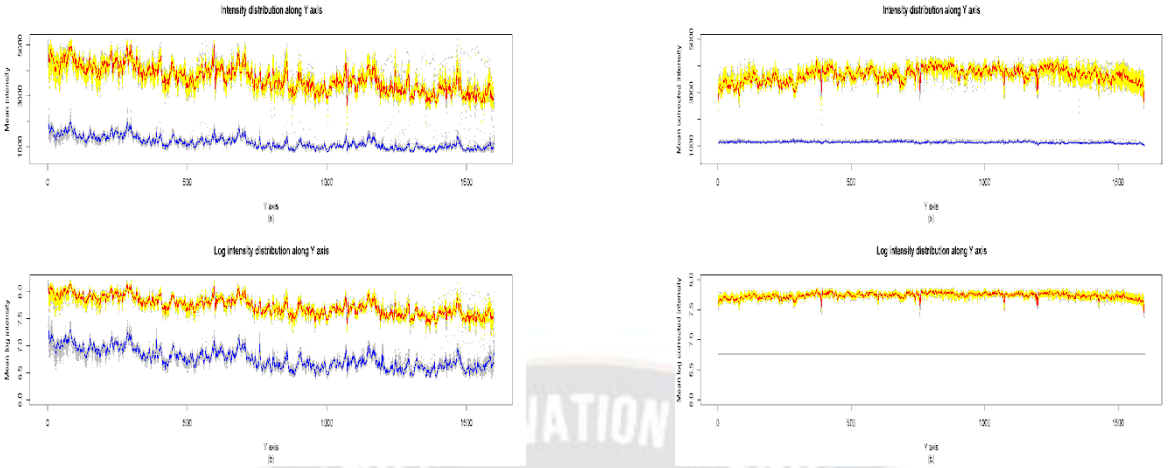

Figure 1. Example of a TWO-COLUMN figure caption: (a) this is the format for referencing parts of a figure.