ISSN: 2633-4828 Vol. 5 No.4, December, 2023

International Journal of Applied Engineering & Technology

ADVANCING HEALTHCARE: HARNESSING MACHINE LEARNING FOR PROTECTING PRIVACY AND HEART DISEASE PREDICTION

Vaseem Ghada and Dr. Gopi Sanghani
Department of Computer Science & Engineering, Darshan University Rajkot, India
vaseem.ghada@darshan.ac.in and gopi.sanghani@darshan.ac.in

ABSTRACT
Protecting privacy when using collaborative association rule learning is crucial and has been applied extensively in medical research. Hospitals create a staggering amount of data at an exponential rate. Processing of this data can facilitate decision-making. Data analytics is vulnerable to privacy abuses, though data analytics greatly aids in decision-making, but there are significant privacy issues that must be addressed. Thus, maintaining individual privacy during the data analytics process becomes a crucial and essential duty. Many numbers of research have been conducted in the past that focus on heart prediction, with particular attention paid to the important characteristics that are crucial in the prediction of heart disease. The right combination of significant factors must be chosen in order to enhance the prediction model's performance. Combining the data from all local hospitals prior to association rule mining would yield more precise results. This research aims to provide insight into the many issues faced by privacy preserving association rule mining (PPARM). Learning and putting into practice the most appropriate strategy for various data scenarios can also be beneficial.

1. INTRODUCTION
Now a days, hospitals have been using electronic health record (EHR) systems more and more, which is a massive amount of recorded patient's data [1]. The association rule mining technique is highly useful in many healthcare applications (e.g., forecasting the chance of disease based on past data, proposing effective therapy, correlation between disease and symptoms, etc.) [2][3]. An EHR system's data collection efforts are crucial to medical research [4]. If association rule mining is carried out and the data from all local EHR systems are merged, the findings will be more accurate. Sharing local EHR data is necessary for the integration of local EHR systems. Patient's right to privacy is violated when medical records are shared. Therefore, the challenge of mining association regulations without disclosing local private EHR data is the main focus of medical research. The issue is resolved through the use of privacy-preserving distributed association rule mining (PPDARM), a technique that extracts association rules while ensuring the confidentiality of patient data.

Cardiovascular illness has consistently been the leading cause of mortality worldwide in recent decades. The proportion of cardiovascular disorders that cause death can be decreased by ensuring perfect diagnosis as well as the provision of suitable treatment. Cardiovascular disease prognosis is based on a multitude of factors. In the past, researchers concentrated more on determining important characteristics to incorporate into their models for predicting heart disease. Determining how these features relate to one another and how important they are within the prediction model were given less weight [5]. Numerous studies linked to data mining have been carried out in the past to address the problems that impede early and accurate diagnosis [6, 7, 8]. Weighted ARM (WARM) is employed for finding the correlations between attributes and identify mining rules that produce specific predictions [9]. Users may easily determine the significance of the characteristics that lead to heart disease by using the weight that is employed in this mining process, which also helps to produce more precise rules [10]. The Associative Rule Mining (ARM) technique can be used to determine the association between each feature that helps predict heart disease [11]. Forecasting cardiac disease also makes use of ARM.

2. DISTRIBUTED ASSOCIATION RULE LEARNING
Distributed association rule learning is a method used in machine learning for mining associations or correlations among a set of items in large datasets that are distributed across multiple computing nodes or systems. Association rule learning aims to discover interesting relationships or patterns in data, typically represented as "if-then" rules. It is characterised as, $I = \{i_1, i_2, i_3 \ldots i_m\}$ represent a collection of unique item sets found in dataset D.

The horizontal partitioning of the dataset D results $D = \{D_1 \cup D_2 \cup D_3 \dots D_n\}$, $D_i \cap D_j = \varphi$, $1 \leq i \neq j \leq n$. The definition of association rule is that, if itemset X occurs in a transaction, then itemset Y is likely to occur in the same transaction as well. Support and confidence in the association rule serve as indicators of its utility and interest. Support measures the frequency or occurrence of a particular itemset in the dataset. It indicates how often a specific combination of items appears together in the transactions relative to the total number of transactions. Mathematically, the support of an itemset X is defined as the ratio of the number of transactions containing X to the total number of transactions in the dataset. Higher support values imply that the itemset occurs frequently in the dataset. Confidence measures the strength of association between two itemset in an association rule. It is defined as the conditional probability that an itemset Y appears in the transactions given that itemset X also appears. In other words, confidence measures how likely it is that itemset Y appears in transactions that contain itemset X. Mathematically, the confidence of a rule X→Y is calculated as the ratio of the support of the combined itemset (X ∪ Y) to the support of the antecedent itemset (X). Higher confidence values indicate a stronger association between the itemset.

2.1 Data Partition Model

A data partition model refers to a method of organizing or splitting a dataset into smaller subsets or partitions for the purpose of analysis or processing. The structure of horizontally partitioned data is identical for all participants; however, the records of various entities are contained in each participant. Figure 1 illustrates how all systems share the same structure yet each local system has unique patient data. Several hospitals are interested for doing ARM in the case of worldwide heart disease research. Every local EHR system has information of patient's disease and other details. A collaborative study aimed at determining the relationship between various attributes using a data. However, hospitals don't not want to share their system data in order to protect patient privacy (per some privacy laws) and the anonymity of local trends. As a result, medical research has given the PPDARM.

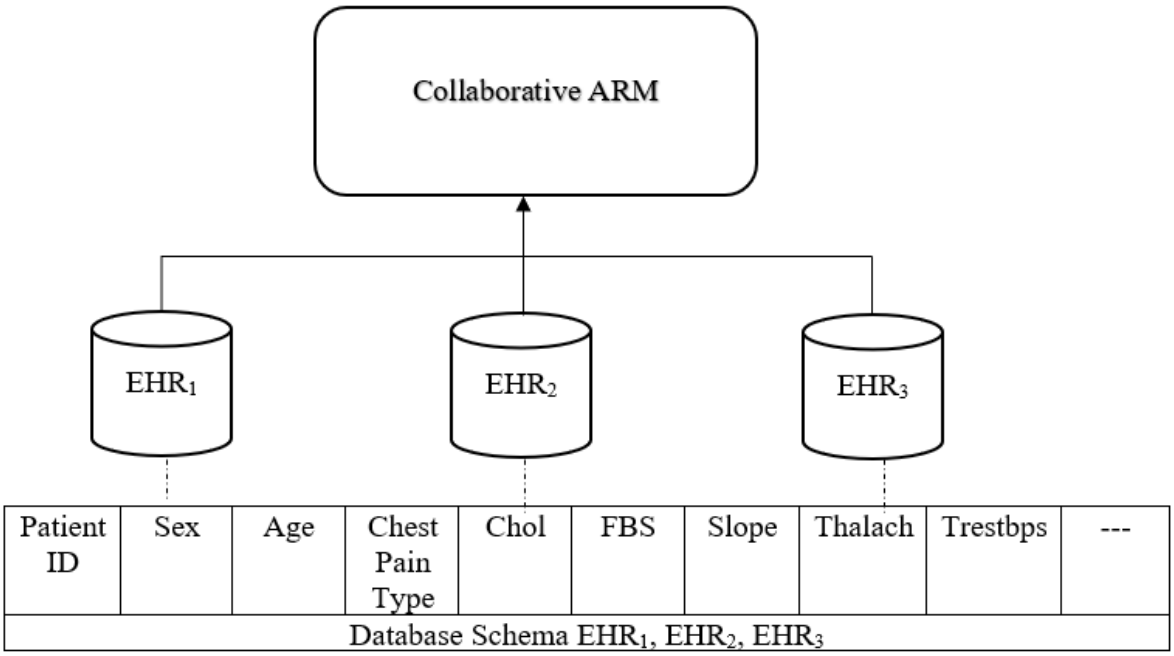

Figure 1: Horizontal partition data model.

Each participant in a vertically partitioned data set has a unique schema, but they all hold the same set of entities' data. Figure 2 depicts the storage of patient data in the hospital, as well as the storage of patient call information by the cell phone company, utilizing a shared identification. Hospitals and mobile companies wish to use association rule mining in medical research to determine how a mobile affects a specific ailment. The hospital

discovers the association rule for instance, that those who use their phones for 18 hours a week are more likely to get brain tumours due to this joint research effort. The data of patients and cell phone users, however, could not be disclosed by the hospital or the phone carrier, respectively

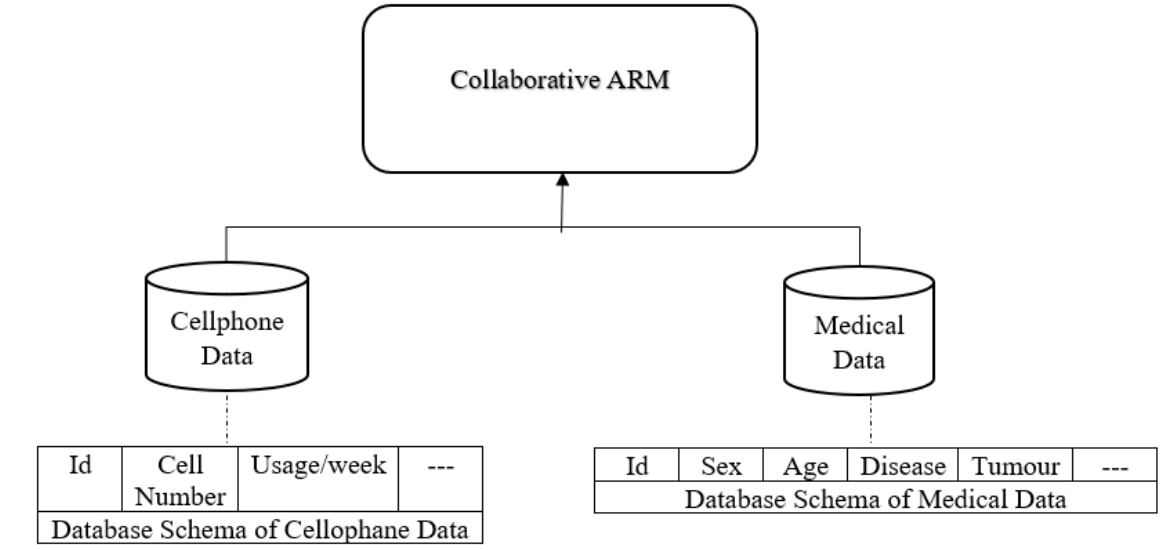

Figure 2: Vertical partition data model.

3. LITERATURE SURVEY OF EXISTING PPARM TECHNIQUES

Numerous algorithms have been created to address the PPARM problem from various angles. As shown in Figure 3, We categorise PPARM using the following four dimensions in this section: (1) Data distribution (2) material that must be safeguarded (3) methods for maintaining privacy and (4) methods for altering data.

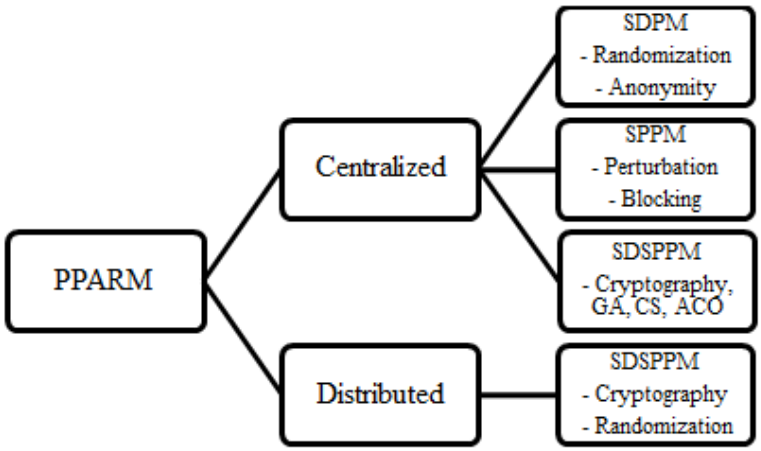

Figure 3: Classification of existing PPARM techniques.

The distribution of data is the first dimension. A few methods for centralised data are suggested. In this case, a sole data owner publishes the data to the external website. Different strategies have been developed for data that is dispersed. Additionally, the dispersed data can be divided into two categories: horizontally distributed data and vertically distributed data. When there is a vertical data distribution, each site has the same quantity of transactions, but each site has a distinct set of attributes. When there are variations in the quantity of transactions at each hospital with the same set of attributes are present across all hospital, this is referred to as horizontal data distribution.

Sensitive raw data or pattern that which requires protection is referred to as the second dimension. While some PPARM algorithms are created to safeguard sensitive patterns, others are created to safeguard sensitive raw data, and yet others are created to safeguard both types of information.

The methods for maintaining privacy when mining association rules are referred known as the third dimension. The six primary categories comprise the developed techniques.

(1) Techniques based on obscure data. These methods, which introduce noise into the original database, can safeguard privacy.

(2) Methods based on heuristics. To reduce the loss of data utility, certain strategies, such adaptive modification, changed only a subset of the values instead of all of them.

(3) Techniques based on reconstruction. These methods do not directly sanitise the source data. These techniques' basic concept is to sanitise the "knowledge base (KB)" before entering the sanitised knowledge base, designated as KB'. Next, the released data is rebuilt using KB'.

(4) Methods based on metaheuristics. These methods are advanced heuristic methods.

(5) Techniques based on cryptography. These methods are frequently applied in situations when association rule mining is delegated to a cloud server. These data that are outsourced can be disseminated or centralised.

(6) SMC stands for Secure Multiparty Computation. SMC is frequently used in distributed databases to accomplish association rule mining. Without disclosing such information to the other parties, SMC aims to extract the global association rules The techniques for altering data are covered in the fourth dimension. To guarantee optimal privacy protection, original data must be modified before a database is made public. A data modification technique is always reliant on the algorithm's chosen privacy-preserving policy. In order to process association rule mining, the original data is modified using the following methods.

(1) Data-obscuring techniques, such data randomization and data anonymity, are employed in association rule mining to safeguard sensitive raw data.

(2) Hiding techniques, such as blocking and perturbation, that are utilised to hide association rules. The term "perturbation" describes the addition of noise or the altering of an attribute value from 1 to 0. Blocking is the process of changing an attribute value that already exists to a question mark (?).

(3) Intelligent metaheuristic-based modification techniques, include the ACO, GA approach, Cuckoo Search (CS) algorithms.

(4) Encryption techniques, like more intricate public-key encryption or substitution encryption.

3.1 Sensitive Data Protection Model (SDPM)

The objective of SDPM is to protect the confidential data stored in the original database. In this paradigm, prior to publishing their data on the external website, the data owner takes measures to sanitize or remove any sensitive material using a range of approaches, with the aim of safeguarding the confidentiality of the raw data. The external site use ARM techniques to do association rule mining on the sanitized database.

Data anonymity and randomization are two well-known techniques for sanitising data that are employed in SDPM. In the data randomization strategy, each sensitive dataset item is assigned a random noise, which is added or multiplied using either the normal distribution or the Gaussian distribution. Additive and multiplicative randomization are two broad categories into which data randomization can be divided. Data randomization techniques are straightforward but effective at concealing sensitive information. Nevertheless, there is a chance that the sanitised dataset records could be utilised to identify a person using public records. One raw data protection technique that circumvents the drawbacks of randomization techniques is data anonymity. The goal of data anonymity is to stop personal records from being identified and disclosed.

3.2 Sensitive Pattern Protection Model (SPPM)

The main goal of SPPM is to prevent the disclosure of its sensitive patterns throughout the association rule mining process on the sanitized database at the external site. The data owner modifies the original database D, also known as the "knowledge base (KB)", to prevent the discovery of specific sensitive association rules using association rule mining techniques, hence protecting the sensitive association rules. Ultimately, the data owner publishes the sanitized database D' to the external website. The sensitive rules Rs are concealed within association rules R', which are produced by the server at the external site following the execution of association rule mining on the sanitized data. The association rule hiding (ARH) model is synonymous with the sensitive pattern protection model. Association rule concealment strategies can be classified into three primary groups: heuristic-based, reconstruction-based, and metaheuristic-based procedures.

(1) Heuristic-based ARH

The most effective method for hiding association rules is to adopt a heuristic approach. Greek verb "heuriskein" which means "to find" is the source of the word "heuristic". Heuristic algorithms, which rely on error-based methods to explore the solution space and find satisfactory solutions, were commonly considered as heuristics. This method typically yields a decent answer, regardless of whether the result can be shown to be accurate. Following two methods are the primary groups of heuristic-based approaches, as per the data modification strategy.

(A) Perturbation-Based Method

The vast majority of scholars who have studied association rule concealing have embraced the perturbation-based approach. The term "perturbation" describes the selective alteration of some attribute values, such as converting a subset of 1 to 0, to lessen the support for rules and maintain the released database's maximum usefulness. These techniques are supposed to focus on the utility issues under the assumption of security. Numerous academic works examined how to reconcile data utility and privacy. Telikani and Shahbahrami's [12] combination of border and heuristic methods allowed for the optimisation of association rule concealment. In this work. For border-based solutions, the MaxMin technique is the best option. Telikani used the MaxMin technique in conjunction with two heuristics to conceal association rules. They suggested the Decrease the Condense of Rule (DCR) hybrid algorithm. Regarding the amount of lost rules, rule mining time, and data utility, the experimental findings demonstrated that DCR performs better than Hai et al.'s proposed ARRHIL algorithm. A side-effect-free ARH approach was proposed by Surendra and Mohan [13]. The unique aspect of the suggested approach is that closed item sets and patterns rather than database transactions are cleaned up. In their scheme, there are lost rules and ghost rules in addition to hiding failure of 0.

(b) Blocking-Based Method

In this method, the sensitive rules are concealed by replacing them with a question mark. for a certain value in a chosen transaction. In certain situations, including in medical applications, it is preferable to use an uncertain value in place of a true one rather than a fake one. The definition of an association rule's support and confidence are somewhat altered when a question mark is added to the dataset. Consequently, the confidence and support will be changed to represent a confidence interval and a support interval, respectively. For association rules that concealed unknown values in the data set, they offered a framework. Aggregation and disaggregation are heuristic-based techniques that are frequently employed in ARH, in addition to perturbation- and blocking-based techniques.

(2) Reconstruction-based ARH

Several solutions were implemented to address the problem of privacy preservation by modifying the data and rebuilding the distributions to achieve association rule concealing. This technique's basic idea is as follows: the data owner finds a so-called "knowledge base (KB)", and uses it to mine the association rules that are hidden in the original database. Next, it sanitises the KB into KB' so that the KB' cannot mine the sensitive rules. Finally, KB' reconstructs the data and releases it to the external website.

(3) Metaheuristic-Based ARH

The PPARM algorithms are anticipated to take into account the utility issues in order to provide security. For conventional PPARH algorithms, minimising side effects during the processing of association rule concealing is a difficult task. It is suggested to use evolutionary algorithms, which are metaheuristic-based solutions, to quickly locate the best answer. Metaheuristics are thought of as high-level ideas for utilising several techniques to explore search spaces. Widely used metaheuristics, or intelligent algorithms with a heuristic framework, include evolutionary algorithms like GA, Cuckoo Search (CS) algorithms, and ACO algorithms. Evolutionary algorithms have been widely used in association rule concealing to find the global optimal solution in recent years. According to the Darwinian principle, GA is a probabilistic search method that uses crossover and mutation to change an original population into a new population known as offsprings. The cpGA2DT [14] technique was presented by Lin et al. to conceal sensitive item sets by eliminating victim transactions that are based on GAs. The authors Lin et al. [14] initially presented a method that uses a multi-objective strategy to hide sensitive data by deleting transactions. Doan et al. [15] implemented ways to mitigate the unintended repercussions of the missing non-sensitive regulations. Their test findings showed that the enhanced strategy performed better. Ant Colony (ACO) algorithms are made to address optimisation problems, just like GA and CS algorithms. ACO-based strategies [16] were put out to enhance performance and lessen adverse consequences.

3.3 Sensitive Data and Sensitive Pattern Protection Model (SDSPPM)

The goal of SDSPPM is to safeguard sensitive patterns as well as sensitive databases. This model is primarily applied to Secure Multiparty Computation (SMC) problems over distributed data and secure outsourcing of association rule mining over vast volumes of data, including distributed and centralised data. Researchers have presented numerous SMC and cryptography-based methods that effectively align with this idea.

(1) Outsourcing of Secure Association Rule Mining

The growth of cloud computing has led to a growing concern over data mining outsourcing. According to our survey, association rule mining outsourcing is primarily concerned with outsourcing related to dispersed or centralised data. The original database D is encrypted by the data owner, who then transfers database D' to the cloud server. After calculating the encrypted strong association rules AR, the miner forwards the results to the data owner. The encrypted rules are decrypted by the data owner, resulting in the generation of the original rules involving the original items. For instance, in order to reduce administrative costs, certain hospitals and organisations frequently outsource mined rule over data to the cloud. The issue of association rule mining over encrypted distributed data has drawn the attention of certain academics in recent years [17].

(2) Secure Association Rule Mining Over Distributed Data

In certain situations, several parties may want to work together to determine the intriguing global association rules over the combination of all partitioned data while keeping their individual data private from one another. It is obvious that standard centralised systems aren't essentially flexible enough for these kinds of circumstances. Privacy Preserving Distributed Association Rule Mining (PPDARM) has become a significant field of study in an effort to address this issue. The primary goal of privacy-preserving association rule mining over vertically distributed data is to securely compute an item set's support count. The miner classifies an itemset as frequent if its support count is greater than or equal to the specified minimum support threshold. Homomorphic encryption was also utilised by Hammami et al. [18] to accomplish privacy-preserving data mining in a cloud computing setting. An attack technique for the symmetric homomorphic encryption scheme was presented by Wang et al. [19]. They demonstrated in [19] that Li et al. overestimated the privacy of their plan. They demonstrated how to use the Euclidean and continuous fraction algorithms to recover the secret key from a number of known plaintext/ciphertext combinations.

4. Dataset

Dataset	UC Irvine Machine Learning Repository
Category	Heart Disease

Selected Database	Cleveland Database
Instances	Total: 310
Features	14
Feature Type	Categorical, Integer, Real
Characteristics	Multivariate
Subject Area	Life Science

Table 1: UCI Dataset

The heart illness data was obtained from the UCI machine learning library. The Cleveland database was selected for this study. The collection contains a total of 303 records. The dataset in the repository contains data for just 14 out of the 76 attributes.

Feature	**Description**	**Value**
Age	Age of the patient in year	Numeric Value
Sex	Gender of the patient	1 for male and 0 for female
Cp	Chest pain type described with 4 values;	Value 1: typical angina, Value 2: atypical angina, Value 3: non-anginal pain, Value 4: asymptomatic
Trestbps	Resting blood pressure	in mm/Hg on admission to the hospital
Chol	Serum cholesterol	in mg/dl
Fbs	Fasting blood sugar >120 mg/dl;	1 if true and 0 if false
Restecg	Resting electrocardiographic results in 3 values;	Value 0: normal Value 1: having ST-T wave abnormality Value 2: showing probable
Thalach	Maximum heart rate achieved	Numeric Value
Exang	Exercise induced angina	1 for yes and 0 for no
Oldpeak	ST depression induced by exercise relative to rest	Numeric Value
Slope	The slope of the peak exercise ST segment	Value 1: upsloping, Value 2: flat, Value 3: down sloping
Ca	Number of major vessels (blood flow disorder)	(0–3) colored by fluoroscopy
Thal	The heart status	Value 3: normal, Value 6: fixed defect, Value 7: reversable
Num	It represents the diagnosis of heart disease.	0 meaning absence, 1–4 meaning presence of disease

Table 2: Description of features of heart disease dataset

The characteristics, its worth, and its explanation are elucidated in Table 2. The prediction of disease involves thirteen features, with one attribute serving as the output or projected attribute for a patient's disease status. The Cleveland dataset includes an attribute named "num" that represents a patient's disease diagnosis. The values of this property vary from 0 to 4. In this scenario, a value of 0 represents the absence of disease, while values ranging from 1 to 4 show the presence of disease in the individuals. The scale reflects the severity of the disease, with 4 indicating the maximum level.

5. WORKFLOW OF IMPLEMENTATION WORK

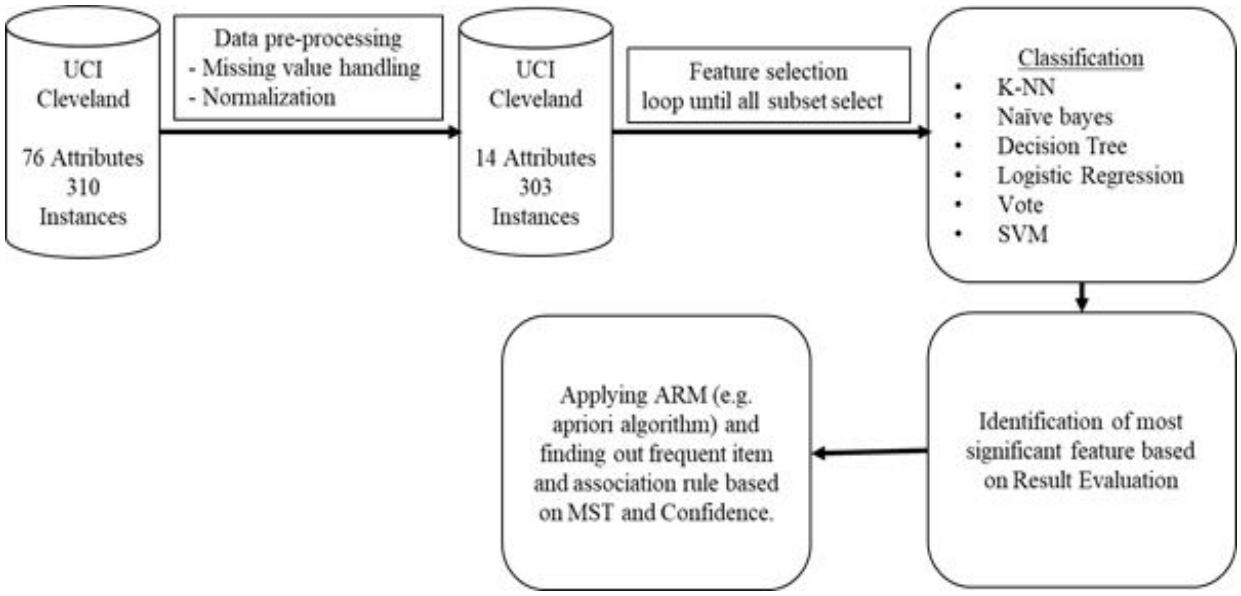

Figure 4: Workflow of Implementation work

Workflow of our implementation work is as shown in figure 4. It's contained total five phase as described below.

5.1 Data Preprocessing

Following data gathering, preprocessing was done. The data had six records with missing values. The dataset's total number of records decreased from 303 to 297 by eliminating all entries containing missing values. Next, the multiclass of the predicted attribute for the presence of heart disease in the dataset were converted to binary values . To complete the data preprocessing operation, all diagnosis values between 2 and 4 were changed to 1. As a result, the only diagnosis values in the generated dataset are 0 and 1, where 0 indicates the absence of heart disease and 1 indicates its existence. Following the reduction and transformation, 160 records for "0" and 137 records for "1" were obtained from the distribution of 297 records for the "num" characteristics.

5.2 Feature Selection

Only the "age" and "sex" characteristics, two of the 13 features, relate to a patient's personal data in the context of heart disease prediction. The clinical characteristics that make up the remaining 11 elements were all gathered from different medical exams. In order to develop the classification model for this experiment, a mixture of features was chosen and employed using different classification techniques: k-NN, Decision Tree, Naïve Bayes, Logistic Regression, Vote, SVM. The brute force approach was used to limit its lower bound (minimum 3 characteristics) for this reason. The process involved testing every potential feature combination using every technique. Using the seven data mining techniques, each of the 13 attributes' possible combinations of three features was investigated in the experiment's initial step. The experiment was then run again to choose every conceivable pairing of four qualities chosen from the list of thirteen attributes.

5.3 Classification Modelling

k-NN, Decision Tree, Naive Bayes, Logistic Regression (LR), SVM, and Vote were the six most widely used classification techniques in data mining that were used to create the models after that feature selection. A 10-folds cross-validation technique was applied to verify the models' performance. Using this method, the dataset is split up into ten subgroups and processed ten times in total. One subset is utilised as a training set, and the other nine are utilised as testing sets. Lastly, an average of the findings after ten iterations is displayed. Because stratified sampling is used to divide the subsets.

5.4 Performance Measure

Three performance metrics were used to assess the effectiveness of the categorization models: precision, f-measure, and accuracy. The fraction of accurately anticipated cases among all instances is known as accuracy. The weighted mean of recall and precision is known as the F-measure. The percentage of accurate predictions for the positive class is known as precision. These three performance measures were used to find the important characteristics; accuracy and precision measures were utilised to find the data mining technique that produced the best-performing models. The three performance indicators offer a more comprehensive comprehension of the overall behaviour of the different feature combinations, which is helpful for the identification of significant characteristics. However, as accuracy and precision are the most logical performance indicators, data mining technique analysis concentrates on the top-performing models that can predict heart disease with high accuracy. Performances have been assessed independently for each classifier, and all data has been accurately documented for future study.

5.5 Result Analysis

The outcomes of the experiments are shown in this section. Based on the examination of the experiment findings, accuracy of the six well known classifiers is measured as shown in the table-3, and from the analysis of combination of selected features, nine significant attributes: sex, cp, fbs, restecg, exang, oldpeak, slope, ca, and thal, were discovered.

Techniques	Accuracy (%)	Combination of selected features
SVM	89.00%	sex, cp, chol, restecg, oldpeak, slope, ca, thal
Vote	84.00%	Age,Sex, cp, thalach, exang, oldpeak, ca
Naïve Bayes	78.04%	Age, sex, cp, chol, fbs, exang, oldpeak, slope, ca,
Logistic Regression	80.48%	Sex, cp, fbs, thalach, exang, slope, ca, thal
k-NN	84.00%	Sex, cp, fbs, restecg, oldpeak, ca, slope
Decision Tree	81.50%	Sex, cp, fbs,chol, oldpeak, ca, thal

Table 3: Accuracy and combination of features of six classifiers.

After selecting nine significant features, accuracy of the five well known algorithms, are measured and the comparison is given in figure 5. As shown in the figure 5, accuracy is increased when we reduce the number of features from 13 to 9.

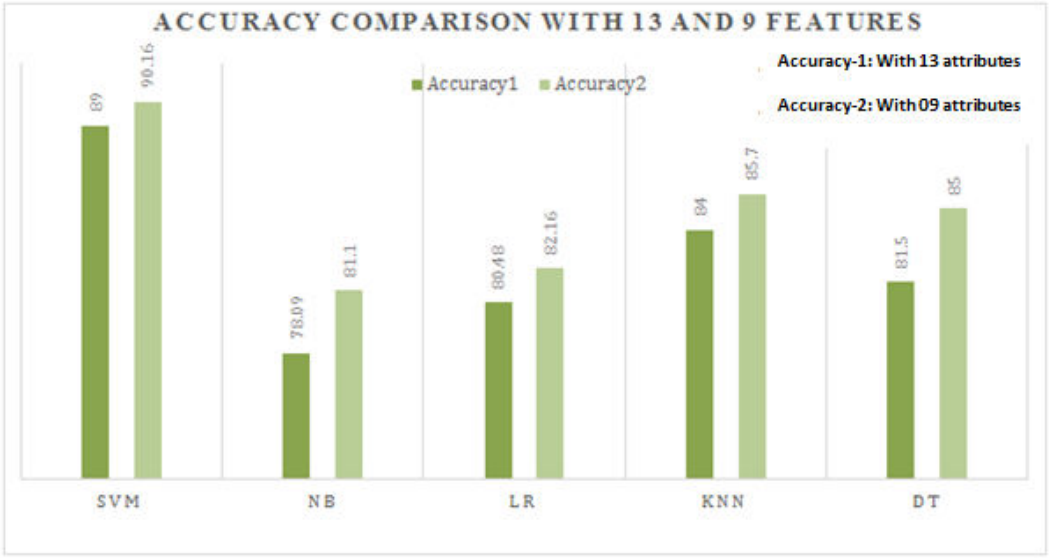

Figure 5: Accuracy comparison with 9 and 13 features .

After selecting the significant features, association rule mining is applied on the data and the frequent item sets, and association rule is generated. Here apriori techniques is applied with the threshold of 50% for minimum support count and 100% of confidence. The generated rule is given as per the table-4 given below.

Rule No	Rule	Support (%)	Confidence (%)
1	{ Thalach, Trestbps, Chol } > Num	51.79	100%
2	{ CA, Chol, OldPeak } > Num	51.07	100%
3	{ Thalach, Sex, Chol } > Num	53.95	100%
4	{ Sex, Chol, OldPeak } > Num	55.39	100%
5	{ OldPeak, Chol } > Num	69.06	100%
6	{ Sex, Chol } > Num	68.34	100%
7	{ Thalach, Chol } > Num	67.67	100%
8	{ Trestbps, Chol } > Num	64.74	100%
9	{ CA, Oldpeak } > Num	61.11	100%
10	{ Thal, Chol } > Num	61.87	100%

Table 4: Generated association rule using apriori algorithm.

7. CONCLUSION

Over the past few decades, PPARM has been extensively studied in the field of health care as a relatively young and quickly developing field of study. There are numerous methods that have been used for PPARM. An exhaustive review of the current PPARM algorithms is given in this study. Our goal in this effort was to collaborate with several EHR systems that include data that is partitioned horizontally in order to increase the accuracy of medical research. Concerns about privacy were brought up in relation to cooperating multiple EHR systems. We suggested a safe and effective PPDARM algorithm for horizontally partitioned data in order to address this. An experimental analysis employing heart disease data confirmed the improved privacy preservation and accuracy of the algorithm.

8. FUTURE WORK

The same experiment can be carried out on a sizable real-world dataset in order to further this topic. Additionally, the suggested method can be expanded to EHR systems with vertically partitioned data and examined for other diseases. As a future work we have planned to implement PPARM using trusted third-party model and semi honest model.

9. REFERENCES

1. Roth GA, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392(10159):1736–88
2. Murphy SL, Xu J, Kochanek KD, Arias E. Mortality in the United States, 2017. NCHS data brief, no 328. Hyattsville, MD: National Centre for Health Statistics;2018.
3. James SL, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet392 (10159), 1789–1858; 2018.
4. Maji S, Arora S. Decision tree algorithms for prediction of heart disease. In Information and communication technology for competitive strategies (pp.447–454). Springer, Singapore; 2019.
5. Mohammed KI, Zaidan AA, Zaidan BB, Albahri OS, Albahri AS, Alsalem MA, Mohsin AH. Novel technique for reorganisation of opinion order to interval levels for solving several instances representing prioritisation in patients with multiple chronic diseases. Compute Methods Programs Biomed. 2020;185:105151.

6. Bashir, S., Khan, Z. S., Khan, F. H., Anjum, A., & Bashir, K. (2019). Improving heart disease prediction using feature selection approaches. In 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST) (pp. 619–623). IEEE.
7. Fitriyani NL, Syafrudin M, Alfian G, Rhee J. HDPM: an effective heart disease prediction model for a clinical decision support system. IEEE Access. 2020; 8:133034–50.
8. Mahdi MA, Al-Janabi S. A novel software to improve healthcare base on predictive analytics and mobile services for cloud data centers, in International conference on big data and networks technologies (pp. 320–339). Springer, Cham; 2019
9. Kannan AG, Castro TARVC, BalaSubramanian R. A comprehensive study on various association rule mining techniques; 2018.
10. Altaf W, Shahbaz M, Guergachi A. Applications of association rule mining in health informatics: a survey. Artif Intell Rev. 2017;47(3):313–40.
11. Chauhan A, Jain A, Sharma P, Deep V. Heart disease prediction using evolutionary rule learning, in 2018 4th International conference on computational intelligence & communication technology (CICT) (pp. 1–4). IEEE; 2018.
12. A. Telikani and A. Shahbahrami, ``Optimizing association rule hiding using combination of border and heuristic approaches," Appl. Intell., vol. 47, no. 2, pp. 544-557, Sep. 2017.
13. H. Surendra and H. S. Mohan, ``Hiding sensitive itemsets without side effects," Appl. Intell., vol. 49, no. 4, pp. 1213-1227, 2018. doi: 10.1007/s10489-018-1329-5.
14. J.C.W. Lin, Y. Zhang, P. Fournier-Viger, Y. Djenouri, and J. Zhang, ``A metaheuristic algorithm for hiding sensitive itemsets," in Proc. Int. Conf. Database Expert Syst. Appl., Regensburg, Germany, 2018, pp. 492-498.
15. K. Doan, M. N. Quang, and B. Le, ``Applied cuckoo algorithm for association rule hiding problem," in Proc. 8th Int. Symp. Inf. Commu. Technol., 2017, pp. 26-33.
16. J. M.-T. Wu, J. Zhan, and J. C.-W. Lin, ``Ant colony system sanitization approach to hiding sensitive itemsets," IEEE Access, vol. 14, pp. 10024-10039, 2017.
17. L. Liu et al., ``Privacy-preserving mining of association rule on outsourced cloud data from multiple parties," in Proc. ACISP, W. Susilo and G. Yang Eds. Cham, Switzerland: Springer, 2018, pp. 431-451.
18. H. Hammami, H. Brahmi, S. Ben Yahia, and I. Brahmi, ``Using homomorphic encryption to compute privacy preserving data mining in a cloud computing environment," in Proc. Eur. Medit., Middle Eastern Conf. Inf. Syst. (EMCIS), Coimbra, Portugal, 2017, pp. 397-413.
19. B. Wang, Y. Zhan, and Z. Zhang, ``Cryptanalysis of a symmetric fully homomorphic encryption scheme," IEEE Trans. Inf. Forensics Security, vol. 13, no. 6, pp. 1460-1467, Jun. 2018.