Journal of Theoretical and Applied Information Technology
15th August 2025. Vol.103. No.15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ROUGH SET BASED PRIVACY PRESERVATION OF HEALTHCARE DATA USING ASSOCIATION RULE MINING AND GENETIC ALGORITHM

SRUTIPRAGYAN SWAIN, BIBHUTI BHUSAN DASH, SUNEETA MOHANTY, BANCHHANIDHI DASH, PRASANT KUMAR PATTNAIK*

[1,3,4,5]School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar, India
[2]School of Computer Applications, KIIT Deemed to be University, Bhubaneswar, India

E-mail: sswain@imit.ac.in ,bdashfca@kiit.ac.in, smohantyfcs@kiit.ac.in
banchhanidhi.dashfcs@kiit.ac.in, patnaikprasantfcs@kiit.ac.in
*Corresponding Author

ABSTRACT

Smart healthcare refers to the combination of emerging trends like Big Data Analytics, IoT and AI & ML. The shift from traditional systems to smart health applications results in the generation of massive volumes of data that need to be managed. This data must be stored on cloud servers and shared with organizations for future research and applications. However, sharing it with third-party servers can result in the exposure of sensitive patient information to external entities. A significant challenge in this context is ensuring the protection of sensitive information to maintain privacy. Consequently, attribute reduction plays a crucial role in managing large datasets by removing unnecessary or redundant data, thereby facilitating the efficient hiding of sensitive rules before public disclosure. This paper presents a privacy-preserving framework designed to hide sensitive fuzzy association rules. The proposed model incorporates two key stages: a pre-activity phase for mining fuzzified association rules and a post-activity phase for concealing sensitive rules. Experimental findings confirm the effectiveness of the proposed approach.

Keywords: *Rough Set (RS), Fuzzy Proximity Ratio (FPR), OIS (Ordered Information System), Fuzzy Association Rule Mining.*

1. INTRODUCTION

Smart healthcare, also known as digital health or health tech, refers to integrating advanced technology with healthcare to improve the efficiency, effectiveness, and personalization of medical services. It leverages advanced technologies to improve the quality, efficiency, and accessibility of healthcare services. E-health records form the basis of storing patient information with the rapid growth in smart healthcare applications. Hospitals and healthcare institutions store voluminous sensitive e-health and Personal health records generated from smart devices. These huge amounts of data are stored in a cloud server shared across third-party organizations for future research on the disease symptoms. In the data-sharing process, sensitive patient data is leaked to the outside world which the patient may not want to disclose. PPDM is the technique of extracting meaningful patterns that are used for future research applications from large datasets without revealing sensitive information about an individual. Real world data contains uncertainty and vagueness. Rough set theory, developed by Zdzisław Pawlak provides an effective mathematical framework for managing the imprecision and uncertainty commonly found in real-world dataset

Association rule mining is the technique of finding the relationship among the attributes of object.

The Genetic Algorithm (GA), originally introduced by Holland, is a widely recognized metaheuristic algorithm aimed at finding optimal solutions within a problem space. Following the principles of natural selection observed in biological evolution, GA employs a sequence of three actions—selection, crossover, and mutation [2] are iteratively applied to evaluate solutions throughout the evolutionary process. The concepts of association rule mining and GA are combined to hide sensitive patient data while releasing a large set of data to the outside world in healthcare applications.

Artificial intelligence encompasses various techniques, one of which is Fuzzy Logic. Fuzzy logic is different from traditional logic by allowing the degree of truth values to range between 0 (false) and 1 (true). It provides a means to represent

concepts with ambiguity, aiding in the organization of ideas within a theoretical framework. Fuzzy Association Rule Mining merges findings from data mining with human insight and prior knowledge to create rules for categorizing data streams. Many fuzzy based models have been proposed in research for privacy protection. The F-Classify [3] is introduced, which employs fuzzy logic for classifying QI & multiple sensitive attributes. The rules evaluate the classes to which the record belongs to. Extensive experiments conducted on healthcare datasets demonstrate that FClassify outperform existing methods in both privacy protection and utility. Additionally, due to its foundation in artificial intelligence, F-Classify exhibits shorter execution times compared to alternative approaches. The research [4] suggests an approach to guarantee personal data confidentiality in cooperative computation during data mining, utilizing an optimization framework. It tackles the confidentiality issue using different approaches, viewing individual privacy as a multi-faceted optimization task. Recognizing the variability of privacy requirements among users, the article introduces a fuzzy optimization approach to accommodate the inherent vagueness of individual privacy. A fuzzy multi-criterion optimization framework is suggested as an extra privacy safeguard, customized for specific privacy considerations. Fuzzy limitations are established according to users' privacy preferences, utilizing the fuzzy set realm to fulfill specific privacy needs in a computational setting. This approach enables data custodians to tailor their confidentiality degrees as required, providing the utmost adaptability. In multi-objective optimization models, each data owner's interests are reflected in identical objective functions but with varying constraints. The research [5] introduces a novel method for preserving sensitive information through the application of fuzzy logic. At first, clustering is performed on the initial dataset, and then noise is applied to the numerical data using a fuzzy membership function, leading to altered data. The clusters generated from the distorted data maintain a relationship with the original clusters, ensuring privacy. In [6], the impact of various fuzzy-based membership functions like PI Shape, Bell Shape, and S Shape on the PPDM method that uses fuzziness to protect sensitive data. In [7], attributes are converted to fuzzy attributes, which prevent the attacker from predicting the exact value thereby preserving privacy, and also the mining activity shows better accuracy. The HMAU [8] method implements an association rule hiding technique by calculating the minimum HMAU value. Genetic Algorithm assesses the suitability of novel structures to choose the optimal population, with suitability determined by a specified cost function. In the context of PPDM challenges, GA has been utilized to optimize the identification of sensitive transactions for deletion, thereby safeguarding privacy. In Genetic Algorithms, a population consists of individuals referred to as chromosomes, each embodying a full solution to a defined problem, often illustrated as a string of binary digits. The chromosome's fitness is established by a variety of factors and methods. Every population contains numerous chromosomes, and the top-performing chromosome is used to generate the subsequent population. The population progresses through subsequent iterations based on survival of fitness. Chromosomes demonstrating adequate fitness are selected for reproduction, whereas those with lower fitness may be chosen less frequently or excluded entirely. Various methods exist, including Roulette-Wheel, Rank, and Tournament selection. A GA based optimized method has been suggested [9] to guarantee the confidentiality of information for the expanding data stored on untrusted third-party cloud storage. GA has been utilized to optimize the identification of sensitive transactions for deletion, thereby safeguarding privacy. Wu et al. [10] devised two sanitization models leveraging GA to obfuscate sensitive details within medical datasets. Both models integrated the concept of multi-thresholds within the fitness function, leading to enhanced concealment of sensitive information. While one model necessitates re-scanning the dataset for each transaction deletion, the other introduces the notion of pre-large items to serve as a buffer in the sanitization process, consequently diminishing the computational overhead for assessment.

Section 2 describes the method and techniques used, Section 3 gives our experimental results, Section 4 discusses our findings, Section 5, and Section 6 offers the conclusion, limitations and future work respectively.

2. METHODS AND TECHNIQUES

The proposed model hides sensitive data before the dataset gets published. Privacy preservation is done by hiding rules that are sensitive. The proposed model is divided into two activities pre activity and post activity, (Figure 1).

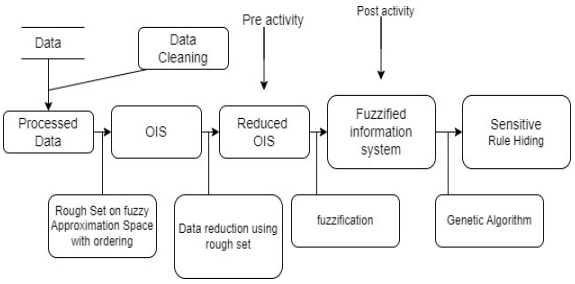

Figure 1: Proposed Model for Privacy Preservation

The real-world data stored in the database are not in discrete format. They are either in qualitative or quantitative format. The first step is to convert them into discrete data. Data cleaning is done first by using RSFAS technique and ordering rules. Then, to process the quantitative or qualitative data, pre activity phase is used. It includes problem understanding, data set, Database cleaning process, FPR, ordering rules and data classification methods For attribute reduction from the information system, the RS reduction technique is applied. Fuzzified association rules are mined by using fuzzification. Further on, Genetic Algorithm is used in post activity to hide sensitive FAR. The main aim of this model is that it will work for both types of databases either qualitative or quantitative data sets or both. Table II presents the proposed sensitive rule hiding model.

The fuzzy proximity relation R (xi, xj) is evaluated using (1)

$$\boldsymbol{R(x_i, x_j) = 1 - \frac{|v_{x_i} - v_{x_j}|}{2(v_{x_i} - v_{x_j})}} \qquad (1)$$

The following contractions are used in our proposed algorithm.

F: Database with fuzzified value.

Mutation to be implemented: Here Bit-flipping is used.

Here single-point cross-over is used.
Here we define the convergence criteria as the maximum count of generations.
Rh: hidden rules.
Xij: ith gene of jth chromosome

n: Region of attributes.
NA – no. of attribute regions
NS –The attributes present in the sensitive rules.

Table 2: Sensitive Rule Hiding Process

Input: *OIS after reduction, Minimum Support, Minimum Confidence.*
Output: *OIS from where nobody can mine the sensitive* FAR.
Step -1 *Reduced Ordered Information System.*
Step -2 *Cleaning of Ordered Information System.*
Step -3 *Fuzzification of reduced OIS using membership function.*
Step -4 *Computation of Support of each item where a ϵ F, F is the database after fuzzification.*
Step -5 *if Min_Supp> all a(support) then*
Exit; // no rule is to be hidden
Step -6 *Search two large item's support values from F;*
Step -7 *for each x's item's support value*
R = {All the rules generate from x}; // where x= {r_1, r_2}, rules can be
$r_1 \rightarrow r_2$, $r_2 \rightarrow r_1$.
7.1 *Calculate confidence of the existing rule is C;*
If Min_Conf< confidence (C)
Then the rule with C is added to Rh;
end if
end for
Step -8 *Extraction of each item in the sensitive rules.*
Step -9 *The transactions are encoded into chromosomes.*
Step -10 *Each chromosome Fitness evaluation with the help of fitness function.*

$$F(Ci) = \sum_{j=1}^{n} \frac{x_{ij}}{support(aj)}$$

Step -11 *The repetition of the steps is done for the specified number of generations*
11.1 *Select the parents with the help of roulette wheel selection.*
11.2 *crossover is performed.*
11.3 *mutation is performed.*
Select: F(Ci) >Min_Conf
11.4 Calculate *f(Curr-gen.)*
*f(curr_gen.) = 0.5 * Difference factor / 100 + 0.5*Modification factor.*
End Repeat.
Step-12 The database F is constructed by changed values in generation with values of minimum fitness.
Step -13 *Output is the OIS where the sensitive rules are hidden.*
Step -14 End

2.1 A CASE STUDY ON PATIENT HEALTHCARE SYSTEM HAVING KIDNEY FAILURE SYMPTOMS

In this section we have demonstrated the proposed model by taking a real-life Kidney failure symptoms dataset. Kidney failure is a perilous condition that is difficult to detect or diagnose during its early stages. The proposed model has been implemented on a real-world dataset of different patients having kidney failure symptoms. The dataset contains the values of different parameters of the kidney failure symptom. The relationship among them is analyzed by computing the FPR. To order the values of different parameters the rough set concept is used followed by fuzzification. The information gathered from ten different Patients is shown in Table 3. The notation Pi is used to denote 10 distinct patients where i=1, 2, 3, . . . ,10.

A. *Pre-activity Process for hiding sensitive rules*

This section processes the proposed model's pre-activity phase. The patient dataset in Table 3 is taken into consideration.

Table 3: Sample information system

Patient's name	**Albumin** g/dL	**BUN** mg/dL	**GFR** mL	**Tot.Chol.** mg/dL	**Triglycer.** mg/dL
P_1	3.2	78.4	15.7	196.5	267.4
P_2	3.3	67.5	20.5	211.8	197.3
P_3	4.1	82.6	18.9	176.3	201.5
P_4	4.5	84.9	23.5	189.5	182.2
P_5	3.7	75.3	14.6	232.7	178.4
P_6	5.8	66.4	14.9	269.7	178.8
P_7	6.7	78.5	30.2	168.4	162.3
P_8	3.6	63.2	19.5	146.8	158.9
P_9	4.8	85.3	25.8	232.5	143.5
P_{10}	5.3	62.8	14.8	167.8	167.9

The FPR (Fuzzy Proximity Relation) is evaluated concerning the attributes. The Parameters that are almost similar to the attribute values are identified, which helps to find the similarity between the objects. The FPR Fr_i =1, 2, 3, 4, 5 for attributes Albumin, Bun, Gfr, Tot. Chol, and Triglyceride are calculated. Table 4 describes the FPR of the attribute Albumin.

Now by considering $\alpha \geq 0.99$, that is the approximate equivalent value of 99% among the objects, it is identified that, R1 (P1, P1) =1; R1(P1, P2) =0.992; R1(P2 , P2) =1; R1(P3 , P3) =1; R1(P4 , P4) =1; R1(P5 , P5) =1; R1(P6,P6) =1; R1(P7, P7) =1; R1(P8 , P8) =1; R1(P9 , P9)=1;R1(P10, P10) =1.Thus, the Patient P1, P2 are α –uniform. Similarly, P3 is α –uniform; P4 is α –uniform; P5 is α –uniform; P6 is α –uniform; P7 is α –uniform; P8 is α –uniform; P9 is α –uniform; P10 is α –uniform. Thus, we get

$U/R1\alpha$={{*P*1,*P2*},{*P*8},{*P*3},{*P*4},{*P*9}},{*P*10},{*P*6},{*P*7},{*P*5}}

Therefore, the values of the attribute Albumin are categorized into nine classes, namely Superb, Excellent, Very High, High, Good, Standard, Bad, Very Bad, Very Very Bad. Now we can order the information system's Albumin attribute. Likewise

for the attributes Bun, Gfr, Tot., Chol, and Triglyceride equivalence classes are generated and represented below.

$U/R2\alpha$={{*P*1,*P*7},{*P*5},{*P*10},{*P*8},{*P*6},{*P*2},{*P*3},{*P*4},{*P*9}}
$U/R3\alpha$={{P1},{P5,P6,P10},{P3},{*P*8},{*P*2},{P4},{P9},{P7}}
$U/R4\alpha$={{*P*1},{*P*4},{*P*3},{*P*7,*P*10},{*P*8},{*P*2},{*P*5},{*P*9},{*P*6}}
$U/R5\alpha$={{*P*1},{*P*3},{*P*2},{*P*4},{*P*5,*P*6},{*P*10},{*P*7} ,{*P*8},{*P*9}

The ordered information system of the kidney failure symptoms to get the patient status of Table 3 is shown in Table 5. Here we consider the weights of Superb, Excellent, Very High, High, Good, Standard, Bad, Very Bad, and Very very Bad as 9,8,7,6,5,4,3,2, and 1 respectively.

$\prec_{Albumin}$:Superb $\prec$Excellent$\prec$Veryhigh $\prec$High $\prec$Good $\prec$ Standard $\prec$Bad $\prec$Very Bad $\prec$ Very very bad
$\prec_{Bun}$:Superb $\prec$Excellent$\prec$Veryhigh $\prec$High $\prec$Good $\prec$ Standard $\prec$Bad $\prec$VeryBad $\prec$Very very bad
$\prec_{Gfr}$:Superb $\prec$Excellent$\prec$Veryhigh $\prec$High $\prec$Good $\prec$ Standard $\prec$Bad $\prec$ Very Bad
$\prec_{Chol}$:Superb$\prec$Excellent$\prec$Veryhigh$\prec$High $\prec$Good $\prec$Standard $\prec$Bad $\prec$VeryBad $\prec$ Very very bad
$\prec_{Triglycer}$:Superb$\prec$Excellent$\prec$Veryhigh$\prec$High$\prec$Good $\prec$Standard $\prec$Bad $\prec$Very Bad $\prec$ Very very bad

Table 4: FPR for the attribute Albumin

R_1	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
P_1	1.000	0.992	0.938	0.915	0.964	0.855	0.823	0.970	0.900	0.876
P_2	0.992	1.000	0.946	0.923	0.971	0.862	0.830	0.981	0.907	0.825
P_3	0.938	0.946	1.000	0.977	0.974	0.914	0.879	0.967	0.961	0.936
P_4	0.915	0.923	0.977	1.000	0.951	0.936	0.901	0.944	0.983	0.959
P_5	0.964	0.971	0.974	0.951	1.000	0.889	0.855	0.849	0.935	0.911
P_6	0.855	0.862	0.914	0.936	0.889	1.000	0.964	0.883	0.952	0.977
P_7	0.823	0.830	0.879	0.901	0.855	0.964	1.000	0.849	0.917	0.942
P_8	0.970	0.981	0.967	0.944	0.849	0.883	0.849	1.000	0.928	0.904
P_9	0.900	0.907	0.961	0.983	0.935	0.952	0.917	0.928	1.000	0.975
P_{10}	0.876	0.825	0.936	0.959	0.911	0.977	0.942	0.904	0.975	1.000

Table 5: OIS for kidney failure symptoms

Patient's name	Albumin g/dL	BUN mg/dL	GFR mL	Tot.Chol.mg/dL	Triglycer.mg/dL
P_1	Very very bad (1)	High (6)	Bad (3)	Good (5)	Superb (9)
P_2	Very very bad (1)	Standard (4)	High (6)	High (6)	Very high (7)
P_3	Standard (4)	Very high (7)	Standard (4)	Bad (3)	Excellent (8)
P_4	Good (5)	Excellent (8)	Very high (7)	Standard (4)	High (6)
P_5	Bad (3)	Good (5)	Very bad (2)	Excellent (8)	Good (5)
P_6	Excellent (8)	Bad (3)	Very bad (2)	Superb (9)	Good (5)
P_7	Superb (9)	High (6)	Superb (9)	Very Bad (2)	Bad (3)
P_8	Very Bad (2)	Vaery Bad (2)	Good (5)	Very very bad (1)	Very bad (2)
P_9	High (6)	Superb (9)	Excellent (8)	Very high (7)	Very very bad (1)
P_{10}	Very high (7)	Very very bad (1)	Very bad (2)	Very bad (2)	Standard (4)

2.2 Post-activity process for hiding sensitive rules

This process aims to find the FAR from the OIS and hide the fuzzified sensitive rules by considering the MS and confidence. Using the triangular membership function, the fuzzified OIS is generated from the OIS is given in Table 6. The

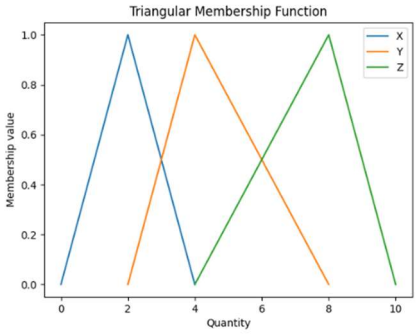

Figure 2: Triangular MF

Table 6: Fuzzified Information System

Patient	A			B			C			D			E		
n	A_x	A_y	A_z	B_x	B_y	B_z	C_x	C_y	C_z	D_x	D_y	D_z	E_x	E_y	E_z
P_1	1	0	0	0	0.5	0.5	0.5	0.5	0	0	0.75	0.25	0	0	0.5
P_2	1	0	0	0	1	0	0	0.5	0.5	0	0.5	0.5	0	0.25	3
P_3	0	1	0	0	0.25	3	0	1	0	0.5	0.5	0	0	0	1
P_4	0	0.7	0.2	0	0	1	0	0.2	3	0	1	0	0	0.5	0.5

		5	5					5							
P5	0.5	0.5	0	0	0.75	0.25	1	0	0	0	0	1	0	0.75	0.25
P6	0	0	1	0.5	0.5	0	1	0	0	0	0	0.5	0	0.75	0.25
P7	0	0	0.5	0	0.5	0.5	0	0	0.5	1	0	0	0.5	0.5	0
P8	1	0	0	1	0	0	0	0.75	0.25	1	0	0	1	0	0
P9	0	0.5	0.5	0	0	0.5	0	0	1	0	0.25	3	1	0	0
P10	0	0.25	3	1	0	0	1	0	0	1	0	0	0	1	0
Count	3.5	3	5.25	2.5	3.5	5.75	3.5	3	5.25	3.5	3	5.25	2.5	3.75	5.5

Considering the MS (minimum support) set to 3.0 and confidence to 80%. Each attribute region's count value is observed. The preset MS value is compared if the value of region count ≥ the preset MS value, then the attribute regions are collected. Here the regions are,

Ax, Ay, Az, By, Bz, Cx,Cy, Cz, Dx, Dy, Dz, Ey and Ez.

Now considering the attribute regions to generate rules and finding the associated confidence value. These selected attribute regions generate many FAR. Some of them are

Ax→Bz, Ax→Cx, Ax→Dx, Ax→Ey, Ay→Bz, Az→Bz, Az→Cz, Az→Dz, Az→Ez, Bz→Az, Cz→Az, Dz→Az, Ez→Az, Ez→Bz, Ez→Cz, Ez→Dz, Bz→Ez, Cz→Ez, Dz→Ez, Bz→Cz, Cz→Dz, Bz→Dz, Cz→Bz, Dz→Cz, Dz→Bz.

The given rule is "Bz → Cz," and its support is calculated using the information provided, in Table 7.

Table 7: Fuzzified Values Of Bz And Cz

Patient	Bz	Cz	Support of (Bz →Cz)
P1	0.5	0	0
P2	0	0.5	0
P3	3	0	0
P4	1	3	1
P5	0.25	0	0
P6	0	0	0
P7	0.5	0.5	0.5
P8	0	0.25	0
P9	0.5	1	0.5
P10	0	0	0
count	5.75	5.25	2

Confidence of Bz→Cz rule is computed as below:

Support count of (Bz→Cz)= 2

Confidence of (Bz→Cz) = 2/5.75 = 34.7 %

This result shows that the rule Bz→Cz is sensitive but need not to be hidden. So, the next sensitive rule is considered which has to be hidden.

Next, considering the rule Ay → Bz and its calculated support is given in Table 8.

Table 8: Support Count Of A_y And B_z

Patient	Ay	Bz	Support of (Ay→Bz)
P1	0	0.5	0
P2	0	0	0
P3	1	3	1
P4	0.75	1	0.75
P5	0.5	0.25	0.25
P6	0	0	0
P7	0	0.5	0
P8	0	0	0
P9	0.5	0.5	0.5
P10	0.25	0	0
count	3	5.75	2.5

Support count (Ay→Bz) = 2.5

Confidence (Ay→Bz) = 2.5/3 = 83.3 %

Hence, this result shows that, the fuzzy association rule Ay→Bz is sensitive and needs to be hidden.

All the rules in which the attributes are considered sensitive are identified, and then these sensitive rules are extracted along with their associated attributes into a new table, which will be referred to as Table IX. For instance,

if the rules Ay→Bz, Ay→Ez, Cy→Bz, Dy→Bz, Cy→Ez, and Dy→Ez are indicated with sensitive, then Table 10 displays the extracted attributes that are involved in the sensitive rules. The initial population is created by

representing each row of Table 9 as a chromosome in a genetic algorithm or an evolutionary algorithm context.

Table 9: Items In The Critical Rule

Patient's Name	Ay	Bz	Cy	Dy	Ez
P_1	0	0.5	0.5	0.75	0.5
P_2	0	0	0.5	0.5	3
P_3	1	3	1	0.5	1
P_4	0.75	1	0.25	1	0.5
P_5	0.5	0.25	0	0	0.25
P_6	0	0	0	0	0.25
P_7	0	0.5	0	0	0
P_8	0	0	0.75	0	0
P_9	0.5	0.5	0	0.25	0
P_{10}	0.25	0	0	0	0
Support	3	5.75	3	3	5.5

1) Genetic Algorithm for Privacy Preservation: The genetic algorithm (GA) is a problem-solving approach that gets inspiration from the natural selection principle. It uses a heuristic search technique. Various optimization and search problems are solved by this technique. In genetic algorithm (GA), offspring (a new population) is generated out of the initial population through two processes known as crossover and mutation. In genetic algorithm, a collection of potential solutions is referred to as a population, and each solution is represented by a chromosome. By utilizing the solution of the old population new population is created. The process of generating new offspring is iteratively repeated until a specific condition or termination criterion is met.

2) Fitness value of the chromosomes: After getting the critical rules the sensitive rule hiding is done by 1) *calculating the chromosome fitness. The fitness evaluation can be done by using the equation (2).*

$$F(Ci) = \sum_{j=1}^{n} \frac{x_{ij}}{support(aj)} \quad (2)$$

Where x– represents the fuzzy value in the chromosome.
n - attribute-regions count.
Each attribute region is represented by aj. The parent is selected considering the fitness of the Chromosomes with better characteristics have an increased probability of being selected. Table 10 presents the fitness values of chromosomes.

Table 10: Fitness Value of The Chromosomes

Patient's Name	F(Ci)
P_1	0.594
P_2	0.878
P_3	1.536
P_4	0.931
P_5	0.2555
P_6	0.045
P_7	0.086
P_8	0.250
P_9	0.336
P_{10}	0.0833

This can be simulated by following algorithm.

3) Apply mutation and crossover:

Step 1. Calculate sum of fitness of all chromosomes in population known as S. It is executed once for every population.
Step 2.r is generated within interval (0, S), where r is a random number.
Step3.Iterate through the population while accumulating the fitness values, ranging from 0 up to the total sum S. Examine the population and calculate the cumulative fitness total from 0 to the sum s. If the cumulative sum s surpasses a certain threshold r, halt the process and provide the corresponding chromosome as the result.
Step 4: Perform Crossover
4.1reiterate the process for the selected chromosomes using step-1 to step-3,
4.2the regions ri and rj are selected.
4.3 if the difference between (ri) and (rj) > =0.5
4.4 perform swapping of (value (ri), value (rj))
4.5 if (value (ri) > value (rj))
4.6 value (ri) = (value (ri) - value (rj)) - 0.5
4.7 if (value (ri) < value (rj))
4.8 value (rj) = (value (rj) - value (ri)) - 0.5
The support values are updated.
End if
Let us take Ay and Dy in Patient p 3 as shown in below Table 11

Table 11: The Fuzzified Values ***Ay , Dy*** *Of P3 Patient*

Patient's Name	Ay	Bz	Cy	Dy	Ez
P3	**1**	3	1	**0.5**	1

Table 12: Modified Values

Patient's Name	Ay	Bz	Cy	Dy	Ez
P_1	0	0.5	0.5	0.75	0.5
P_2	0	0	0.5	0.5	3
P3	**0.5**	3	1	**0**	1
P4	0.75	**0.0**	0.25	1	0.5
P_5	0.5	0.25	0	0	0.25
P_6	0	0	0	0	0.25
P_7	0	0.5	0	0	0
P_8	0	0	0.75	0	0
P_9	0.5	0.5	0	0.25	0
P_{10}	0.25	0	0	0	0
Support	**2.5**	**4.75**	3	**2.5**	5.5

4) Fitness of the generation: Fitness of the generation
is calculated as follows:
Fitness =0.5 * Difference factor/100 + 0.5 *
Modification factor
A weight factor of 0.5 is assigned for each of the above factors.

$$Difference\ factor = \frac{\sum_{i=1}^{n} diff\ i}{n} \qquad (2)$$

Where, difference is calculated by subtracting the new_confidence from old_confidence.n represents the number of rules i.e. of interest. old_confand new_conf are the rule's confidence values before and after genetic operation respectively.The difference of all the rules is calculated in the below Table 13.

Value (Ay) – Value (Dy) = 1.0 – 0.5 = 0.5 ≥ 0.5 So, swap the values of the attributes. In the next step, the higher value attribute is substituted by a reduced value of (- 0.5), i.e., Dy = (1.0-0.5) - 0.5 = 0.5 - 0.5 = 0.
After that, the child's fitness is calculated. If the value is more than the parent then mutation is performed based on mutation probability. Then the parent is replaced by the new individual.
After that the new single chromosome undergoes mutation. A random number is generated When the probability of mutation is set at 20% of the population size. Otherwise, if it is <0.2 then the mutation of the new chromosome is performed as follows:
Step-5 The loop is repeated until the desired no. of mutation count is achieved.
If the value of support is greater than MS
The fuzzy value of 1.0 is changed to 0.0.
From Table 8 it can be inferred that the Bz region's support i.e., 5.75 is greater than MS.
In Table 8 region Bz has the support 5.75 >MS. The value 1.0 in P4 of region Bz, is changed to 0.0 by mutation. Table12 shows all the modified values after mutation and crossover.

Table 13: Difference Calculation Table

Sl.no	Rules	Old_confidence (%)	New_confidence(%)	Difference
1	Ay→Bz	83.33	50.00	33.33
2	Ay→Dy	50.00	40.00	10.00
3	Cy→Bz	58.33	50.00	8.33
4	Dy→Bz	75.00	30.00	45.00
5	Dy→Ez	66.66	60.00	6.66
6	Bz→Dy	39.13	15.78	23.35

Difference factor = (33.33 + 10.00 + 8.33 + 45.00 + 6.66+ 23.35) / 6 = 21.11
Modification factor =Total Number. of attribute-regions that are modified / Number of attribute regions
Modification factor = 3/15 = 0.2
Fitness = 0.105+ 0.1 =0.205
We repeat the steps mentioned above iteratively until we find the required number of generations. The values which are modified of the generation with fitness values that are least in value are substituted with the previous fuzzified values of the fuzzification information system. After computation, it can be observed that the confidence is 50.00%. So, we can state that the sensitive rule, which is not hidden after fuzzification now it's hidden completely. The remodeled OIS with hiding sensitive rule Ay→Bz and the OIS after

defuzzification are shown in the Table 14 and Table 15 respectively.

Table 14: Fuzzified Information System after modification

Patient.	A			B			C			D			E		
n	A_x	A_y	A_z	B_x	B_y	B_z	C_x	C_y	C_z	D_x	D_y	D_z	E_x	E_y	E_z
***P*1**	1	0	0	0	0.5	0.5	0.5	0.5	0	0	0.75	0.25	0	0	0.5
***P*2**	1	0	0	0	1	0	0	0.5	0.5	0	0.5	0.5	0	0.25	3
***P*3**	0	**0.5**	0	0	0.25	3	0	1	0	0.5	**0**	0	0	0	1
***P*4**	0	0.75	0.25	0	0	**0.0**	0	0.25	3	0	1	0	0	0.5	0.5
***P*5**	0.5	0.5	0	0	0.75	0.25	1	0	0	0	0	1	0	0.75	0.25
***P*6**	0	0	1	0.5	0.5	0	1	0	0	0	0	0.5	0	0.75	0.25
***P*7**	0	0	0.5	0	0.5	0.5	0	0	0.5	1	0	0	0.5	0.5	0
***P*8**	1	0	0	1	0	0	0	0.75	0.25	1	0	0	1	0	0
***P*9**	0	0.5	0.5	0	0	0.5	0	0	1	0	0.25	3	1	0	0
***P*10**	0	0.25	3	1	0	0	1	0	0	1	0	0	0	1	0
Count	3.5	**2.5**	5.25	2.5	3.5	**4.75**	3.5	3	5.25	3.5	**2.5**	5.25	2.5	3.75	5.5

Table 14: OIS in which sensitive rule Ay→Bz *is hidden*

Patient's name	**Albumin g/dL**	**BUN mg/dL**	**GFR mL**	**Tot. Chol mg/dL**	**Triglycerides mg/dL**
P_1	Very very bad (1)	High (6)	Bad (3)	Good (5)	Superb (9)
P_2	Very very bad (1)	Standard (4)	High (6)	High (6)	Very high (7)
P_3	Standard (4)	Very high (7)	Standard (4)	Bad (3)	Excellent (8)
P_4	Good (5)	**0**	Very high (7)	Standard (4)	High (6)
P_5	Bad (3)	Good (5)	Very bad (2)	Excellent (8)	Good (5)
P_6	Excellent (8)	Bad (3)	Very bad (2)	Superb (9)	Good (5)
P_7	Superb (9)	High (6)	Superb (9)	Very Bad (2)	Bad (3)
P_8	Very Bad (2)	Very Bad (2)	Good (5)	Very very bad (1)	Very bad (2)
P_9	High (6)	Superb (9)	Excellent (8)	Very high (7)	Very very bad (1)
P_{10}	Very high (7)	Very very bad (1)	Very bad (2)	Very bad (2)	Standard (4)

3. DISCUSSION

The experiment was carried out by using Windows OS on a real-world dataset of 10 patients using Python. From Fig.3 it can be concluded that our method hides sensitive rules effectively compared to the HMAU method. The effectiveness of the postulated algorithm is determined by the help of parameters such as support; count and the number of rules hidden by considering different MS. The utility is lost if more rules are hidden completely. So careful consideration must be done while considering the rules to be hidden. Our model takes different factors like support and count into consideration for the effective determination of sensitive rules. It can be concluded that our method hides sensitive rules effectively as the number of rules increases. Fig.4 depicts that the system is stable after going through some generations. Our approach minimizes the data redundancy in the pre activity phase and the fuzzified sensitive rules are successfully hidden in the post-activity part.

The model is divided into two phases such as pre activity and post activity. The pre activity phase is used for the extraction of fuzzified association rules. The post activity is used to hide the sensitive rules by using genetic algorithm. The number of non-sensitive rules are maximized from the modified dataset by minimizing the number of modifications to the data. Further, the side effects have been reduced while selecting a particular generation with minimum modification. Further, many research have been conducted to preserve privacy. However, there are drawbacks to find out the efficient solution. Accuracy is one of the parameter must be considered while considering such techniques. So, optimization of the model may be considered for better accuracy and efficiency.

4. COMPARATIVE ANALYSIS

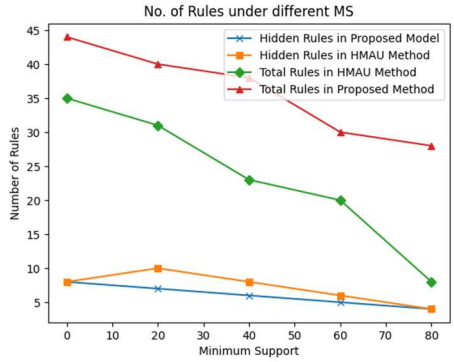

Figure 3: No. of rules under different MS

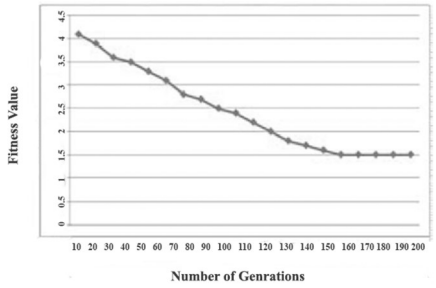

Figure 4: Fitness graph of sensitive rules

4. CONCLUSION

Privacy preservation has emerged as a challenging issue in today's information society. In this study, a model to hide sensitive rules of a patient dataset has been proposed before the data gets published. Our model has two activities pre activity and post activity. We use the pre activity to fuzzified association rules and post activity to hide sensitive FAR. A case study on kidney failure symptoms has been taken into consideration and the sensitive rule hiding model was applied to it. The experiment result shows that the model successfully hides the sensitive rule effectively.

5. FUTURE WORK

The sensitive rules are completely hidden effective utilization of data may not be possible in future research applications. So, finding techniques that not only hide sensitive data but also preserve utility will be our future goal. In addition the model will be optimized by applying firefly algorithm.

REFERENCES:

[1] J Dansana, MR Kabat, PK Pattnaik, "A Novel Optimized Perturbation-Based Machine Learning for Preserving Privacy in Medical Data", *Wireless Personal Communications, Springer*, Vol 130, No.3,2023, pp.1905-1927.

[2] L. Zhang, W. Wang and Y. Zhang, "Privacy preserving association rule mining: Taxonomy, techniques, and metrics", *IEEE Access*, Vol.7,2019, pp.45032-45047.

[3] H.Attaullah, A.Anjum, T.Kanwal, S.U Malik, A.Asheralieva, H.Malik, ... & M.A. Imran," F-classify: Fuzzy rule based classification method for privacy preservation of multiple sensitive attributes",*Sensors*,Vol.21,No.14,2021,pp.4933.

[4] J.Dansana, MR Kabat and PK Patnaik, "Improved 3D Rotation-based Geometric Data Perturbation Based on Medical 10 Data Preservation in Big Data", *International Journal of Advanced Computer Science and Applications*, Vol 14, No.5,2023.

[5] S.Ahmed, S.Haque and S.F. Tauhid," A fuzzy based approach for privacy preserving clustering," *Int. J. Sci. Eng. Res*,Vol.5,no.2,pp. 1067-1071.

[6] S.S.Rathna, and T.Karthikeyan ," Fuzzy membership functions in privacy preserving data mining", *International Journal of Current Research*, Vol.8, no.03,pp.28325-28329.

[7] M.Sridhar, and B. R. Babu, "A fuzzy approach for privacy preserving in data mining", *International Journal of Computer Applications*, Vol.57,no.18.

[8] S.Swain, P. K. Pattnaik, and B.Dash," A Review on Privacy Preservation in Cloud Computing and Recent Trends," in *ICMETE*,SRM Delhi, India, September 22-23, 2023,pp. 365-376

[9]M. Kamal, S. Amin,F.Ferooz , M.J Awan , MA Mohammed , O. Al-Boridi and K.H. Abdulkareem," Privacy-aware genetic algorithm based data security framework for distributed cloud storage," *Microprocessors and Microsystems*,Vol.94,2022,pp. 104673.

[10] J.T Wu, G Srivastava, A Jolfaei, P Fournier-Viger and JCW L," in Hiding sensitive information in eHealth datasets," *Future Generation Computer Systems*,vol.117,2021,pp. 169-180.

.