ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 11, November 2017

Enhancing the Privacy Preserving Data Mining Techniques by using Anatomisation with Slicing Approach for Multiple Susceptible Attributes without Loss of Accuracy

[1] V.K.Saxena, [2] Shashank Pushkar
[1] Vikram University, Ujjain, M.P., India
[2] Birla Institute of Technology, Mesra, Ranchi, Jharkhand, India

Abstract:- **A gigantic quantity of individual health information is accessible in modern decades and dispositioning of any part of this information establishes a huge risk in the field of health care. Enduring anonymization methods are only appropriate for single susceptible and low down dimensional data to remain with privacy particularly like generalization and bucketization. We propose an anonymization technique that is a amalgamation of the betterment of anatomization, and improved slicing approach observing to the principle of k-anonymity and l-diversity for the reason of dealing with high dimensional data along with multiple susceptible data. The anatomization approach disrupts the correlation detected between the quasi identifier attributes and susceptible attributes (SA) and turnouts' two different tables with non-overlapping attributes. Hence, experimental outcomes specify that the suggested method can preserve privacy of data with various sensitive attributes. The anatomization approach reduces the loss of information and slicing algorithm advices in the correlation preservation and usefulness which gives output in sinking the data dimensionality and information deficiency.**

Keywords:- **Privacy preservation, Anatomization, Slicing, k-Anonymity, l-Diversity**

I. INTRODUCTION

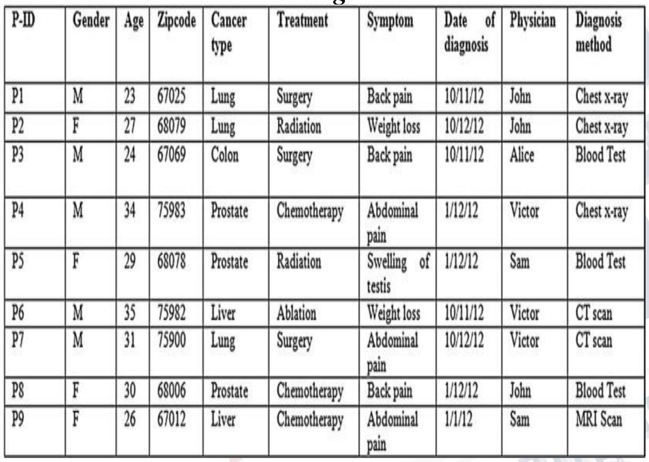
Now-a-days in the field of health care, electronic health record spreads a vast amount of susceptible data in their business. The susceptible data can be personally identifiable information from the customers. Any kind of mishandling of this information creates a crucial warning to their business. When making the susceptible data available to the community, it is necessary for them to shield it from any mistreatment. For data privacy protection, data anonymization is the one and only widely used approach. It customizes information, keeping in mind to make it complicated to link individuals with their data. This methodology tries to guarantee the identity along with the susceptible information of the data subjects when data is shared for varied purposes (LeFevre et al. 2008; Aggarwal et al. 2005; Pfitzmann and Hansen 2008). Sensitive attributes is the set of attributes whose values are undisclosed such as cancer type, treatment, symptom, date of diagnosis and doctor. When sharing records, it is very important to avoid the disclosure of susceptible information of the persons. Anatomization and permutation dissociate the correlation between QI attributes and sensitive attributes by collection and rescheduling of susceptible values in a *qid* group. Slicing is a technique that can undertake with high dimensional data and hence preserve privacy and get better utility (Shyamala and Christopher 2016).

II. CONTRIBUTION

This approach combines the betterments of both anatomisation and improved slicing algorithm following to the theory of l-diversity and k-anonymity. Hence it accords with the complexity in conducting the multiple sensitive attributes in elevated dimensional data. For instance, in Table 1, six susceptible attributes is considered. The attributes showed in the table are Patient-Id, Gender, Zip code, Age, Cancer type, Treatment, Symptom, Date of diagnosis, Physician, Diagnosis method. Out of these attributes, Cancer type, Treatment, Symptom, Date of diagnosis, Physician and Diagnosis method are the sensitive attributes. On the other hand, QI attributes comprises of

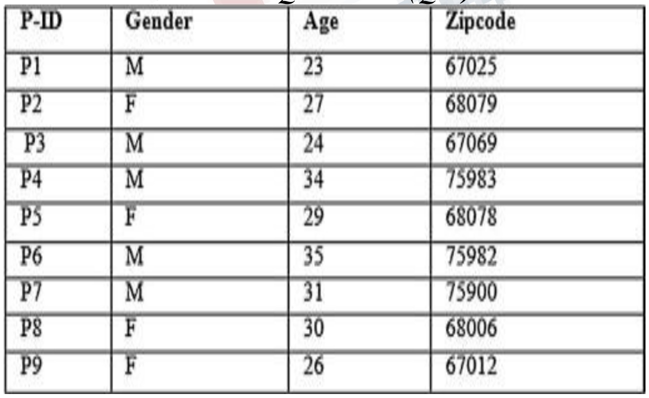
Patient-Id, Gender, Zip code, Age. Basically, the suggested approach anatomizes Table 1 by disconnecting the QI attributes from sensitive attributes and provides two tables, one for the QI attributes and the other for the SA. The outcomes are shown in Tables 2 and 3. This is consequently pursued by employing a slicing technique to Tables 2 and 3. This isolates the tables together horizontally and vertically. In the vertical partitioning stage, the superior clustering algorithm is applied to the Table 3 and the resultant of this step are vastly correlated attributes are in one column. For example {Cancer type, Treatment}, {Symptom, Date of diagnosis}, {Physician, Diagnosis method} are the correlated attributes. Hence, Table 3 is partitioned into three different sensitive attribute tables. In an analogous way the QI attributes in Table 2 are partitioned such that extremely correlated attributes are in one column. For example {{Gender, Age}, {Zip code}}.

Table 1. Original table



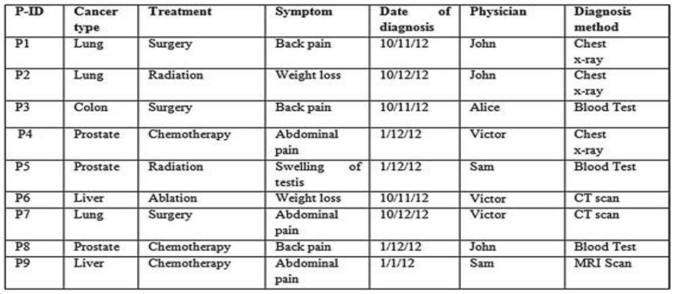
P-ID	Gender	Age	Zipcode	Cancer type	Treatment	Symptom	Date of diagnosis	Physician	Diagnosis method
P1	M	23	67025	Lung	Surgery	Back pain	10/11/12	John	Chest x-ray
P2	F	27	68079	Lung	Radiation	Weight loss	10/12/12	John	Chest x-ray
P3	M	24	67069	Colon	Surgery	Back pain	10/11/12	Alice	Blood Test
P4	M	34	75983	Prostate	Chemotherapy	Abdominal pain	1/12/12	Victor	Chest x-ray
P5	F	29	68078	Prostate	Radiation	Swelling of testis	1/12/12	Sam	Blood Test
P6	M	35	75982	Liver	Ablation	Weight loss	10/11/12	Victor	CT scan
P7	M	31	75900	Lung	Surgery	Abdominal pain	10/12/12	Victor	CT scan
P8	F	30	68006	Prostate	Chemotherapy	Back pain	1/12/12	John	Blood Test
P9	F	26	67012	Liver	Chemotherapy	Abdominal pain	1/1/12	Sam	MRI Scan

Table 2. Quasi table (QIT)



P-ID	Gender	Age	Zipcode
P1	M	23	67025
P2	F	27	68079
P3	M	24	67069
P4	M	34	75983
P5	F	29	68078
P6	M	35	75982
P7	M	31	75900
P8	F	30	68006
P9	F	26	67012

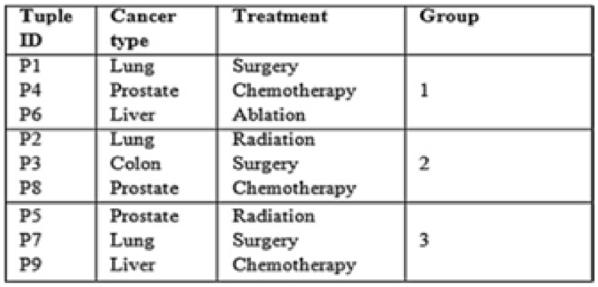
Table 3. Sensitive attribute table (ST)



P-ID	Cancer type	Treatment	Symptom	Date of diagnosis	Physician	Diagnosis method
P1	Lung	Surgery	Back pain	10/11/12	John	Chest x-ray
P2	Lung	Radiation	Weight loss	10/12/12	John	Chest x-ray
P3	Colon	Surgery	Back pain	10/11/12	Alice	Blood Test
P4	Prostate	Chemotherapy	Abdominal pain	1/12/12	Victor	Chest x-ray
P5	Prostate	Radiation	Swelling of testis	1/12/12	Sam	Blood Test
P6	Liver	Ablation	Weight loss	10/11/12	Victor	CT scan
P7	Lung	Surgery	Abdominal pain	10/12/12	Victor	CT scan
P8	Prostate	Chemotherapy	Back pain	1/12/12	John	Blood Test
P9	Liver	Chemotherapy	Abdominal pain	1/1/12	Sam	MRI Scan

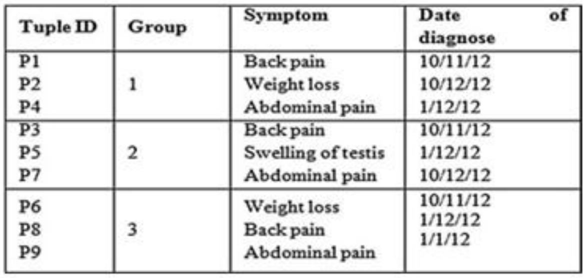
The horizontal partitioning phase is referred to as tuple partitioning. In this phase, tuples in each of the 3 sensitive attributes are bucketised making use of MFA. In order to attain l- diversity, attributes in each bucket are chosen as a distinct one. As the tuple imposes 3 diversity, each bucket also imposes 3 different attributes and this is seen in Tables 4, 5 and 6. For example in Disease column, the different SA {Lung, Prostate, Liver} are in bucket 1, {Lung, Colon, Prostate} are in bucket 2 and {Prostate, Lung, Liver} are in bucket 3. Similarly, the tuples in QIT are bucketised making use of MFA. In order to attain 3 anonymity, for all the individuals in Table 2, there are at least 3 individuals that are con-nected to the same bucket of sensitive values. The bucket imposes 3 different attributes,

Table 4. Sliced sensitive attributes (cancer-treatment)



Tuple ID	Cancer type	Treatment	Group
P1 P4 P6	Lung Prostate Liver	Surgery Chemotherapy Ablation	1
P2 P3 P8	Lung Colon Prostate	Radiation Surgery Chemotherapy	2
P5 P7 P9	Prostate Lung Liver	Radiation Surgery Chemotherapy	3

Table 5. Sliced sensitive attributes (Symptom- date)



Tuple ID	Group	Symptom	Date of diagnose
P1 P2 P4	1	Back pain Weight loss Abdominal pain	10/11/12 10/12/12 1/12/12
P3 P5 P7	2	Back pain Swelling of testis Abdominal pain	10/11/12 1/12/12 10/12/12
P6 P8 P9	3	Weight loss Back pain Abdominal pain	10/11/12 1/12/12 1/1/12

Table 6. Sliced sensitive attributes (physician-diagnosis method)

Tuple ID	Group	Physician	Diagnosis method
P1	1	John	Chest x-ray
P3		Alice	Blood Test
P6		Victor	CT scan
P2	2	John	Chest x-ray
P5		Sam	Blood Test
P7		Victor	CT scan
P4	3	Victor	Chest x-ray
P8		John	Blood Test
P9		Sam	MRI Scan

and this is seen in Table 7. For example, in {Age, Sex} column the different attributes are {(23, M)}, {(24, M)}, {(26, F)}. All the tables from Tables 4, 5, 6 and 7 have a common column referred to as group Id for linking. This Id does the mapping of the QIT with the multiple ST. In this way, the horizontal and vertical partitioning aids in the elimination of the dimensionality of the dataset. For example, in the first bucket of the sliced table as indicated in Table 8 the values are then permutated stochastically in such a manner that the linkage observed between the two columns in one bucket gets hidden. Consider the patient Id P2 with QI values (f, 27, and 68079). In order to decide P2's multiple sensitive values, P2's matching bucket has to be decided. By examining (f, 27), it is a known fact that P2 should be present in bucket 2, as there seems to have no matches in bucket 1 and bucket 3. Then by examining the Zip code attribute in bucket 2, the matching value is (68079, 2, 1, 2). This indicates that {Physician—Diagnosis} method points to bucket 2 of Table 6, {Symptom—Date} points to bucket 1 of Table 5.

Table 7. Sliced quasi identifier attributes

Tuple ID	Gender, Age	Zipcode	Group id
P1	(23,M)	67025	1
P3	(24,M)	67069	1
P9	(26,F)	67012	1
P2	(27,F)	68079	2
P5	(29,F)	68078	2
P8	(30,F)	68006	2
P4	(34,M)	75983	3
P6	(35,M)	75982	3
P7	(31,M)	75900	3

Table 8. Anonymised data

(Gender,age)	(Zipcode, Physician - Diagnosis method, Symptom – Date,Cancer – Treatment)
(23,M)	(67069,1,2,2)
(24,M)	(67012,3,3,3)
(26,F)	(67025, 1,1,1)
(27,F)	(68078,2,2,3)
(29,F)	(68006,3,3,2)
(30,F)	(68079,2,1,2)
(34,M)	(75982,1,3,1)
(35,M)	(75900,2,3,3)
(31,M)	(75983,3,1,1)

and {Cancer—Treatment} points to bucket 2 of Table 4. It is inferred that (f, 27, 68079) may have different values like{John, Sam, Victor} for Physician and,{Chest X-ray, Blood Test, CT scan}for diagnosis method, {Back pain, Weight loss, Abdominal pain}for Symptom, {10/11/12, 10/12/12, 1/12/12} for date of diagnosis and {Lung, Colon, Prostate}for cancer type and {Radiation, Surgery, Chemotherapy} for treatment. Thus, it results in 3 anonymity, because when an individual is mapped onto some sensitive value, at least 2 other individuals are also mapped to the same sensitive values. And it satisfies 3 diversity because it poses 3 distinct sensitive values in each bucket. Thus release of the QI values preserve privacy such that the sensitive value pertaining to an individual that is involved in the QIT can be rightly guessed by an intruder with the computed probability of at most 1/3. The anatomization approach reduces the loss of information through the direct release of the QI attributes.

III. BACKGROUND

Privacy preserving data mining (PPDM) is a rapidly growing research area aiming at eliminating privacy breaches which may happen during the mining of data (Verykios et al. 2004; Kantarcioglu et al. 2004; Clifton 2009). The goal of PPDM algorithm is to alter the original data for the purpose of maintaining privacy, leading to a low degree of data leakage. The work introduced in (Verykios et al. 2004) observes the PPDM approach in the light of five different dimensions. They are data distribution, data modification, data mining algorithm, data or rule hiding and privacy preservation. The work introduced in (Friedman et al. 2008) yields the possibilities for the construction of k-anonymous data models with k-anonymous data sets. More commonly, the k-anonymity concept is utilized by the PPDM algorithms in order to guarantee privacy (El Emam and Dankar 2008). It is a problem to be able to find optimal k-anonymous datasets through generalization and is rated as NP-Hard (Gedik and Liu 2008; Meyerson and Williams 2004). The work indicated in (Li and Li 2006) compares the general taxonomy and multiple generalization schemes. The work that is shown in (Iyengar 2002) formulated a genetic framework in order to look for the best set of generalization for the purpose of satisfying k-anonymity constraints.

Anatomization (Xiao and Tao 2006) in contrast to generalization and suppression does not make modifications to the QI or the SA. The greatest benefit of anatomy is that there is no modification of data in both QIT and ST. Tao et al. (2009) proposed an approach referred to as a permutation, sharing the same kind of spirit of anatomization. A new slicing (Li et al. 2012) approach is employed, which partitions the attributes both horizontally and vertically and avoids membership disclosure. Slicing with Modified Fully Self Adaptive Resonance Neural Network and Metaheuristic Fireflies with Minkowsi Distance Measure (Shyamala and Christopher 2015) improved the clustering accuracy in privacy preservation. But all of the above methods suit single sensitive data only. In order to deal with multiple SA a multiple sensitive bucketization (MSB) (Yang et al. 2008) is suggested. But it is appropriate for attributes less than three only. In case of more SA slicing with MSB.KACA (Han et al. 2013) is brought into use.

Inspired by these works, this paper aims at preserving the privacy of data with multiple SA. In this paper, an anonymization technique is proposed that is a combination of the benefits of anatomization, and enhanced slicing approach adhering to the principle of k-anonymity and *l*-diversity for the purpose of dealing with high dimensional data along with multiple sensitive data.

IV. PRELIMINARIES

4.1. Definition: Anatomization

Anatomy dissociates the correlation between QI attributes with sensitive attributes and generates QIT and sensitive table. QIT has the dataset (q1, q2, . . . , qn) and sensitive table has the dataset (s1, s2, . . . , sm). The idea behind the anatomy approach is that in case two tables with a join attribute goes for publishing, then join corresponding to the two tables could be lossy and again this lossy join is useful in concealing the private data (Xiao and Tao 2006).

4.2. Privacy requirement

The privacy requirement required for publishing multiple sensitive data are k-anonymity and *l*-diversity. K-anonymity (El Emam and Dankar 2008), that helps in preventing the individual records identification in the data, and *l*-diversity (Machanavajjhala et al. 2006), which, on the other side, avoids the association of an individual record having a SA value. In this approach slicing with k-anonymity guarantees that when an individual is mapped onto some sensitive values, at least k − 1 other individuals are also mapped onto the same sensitive values.

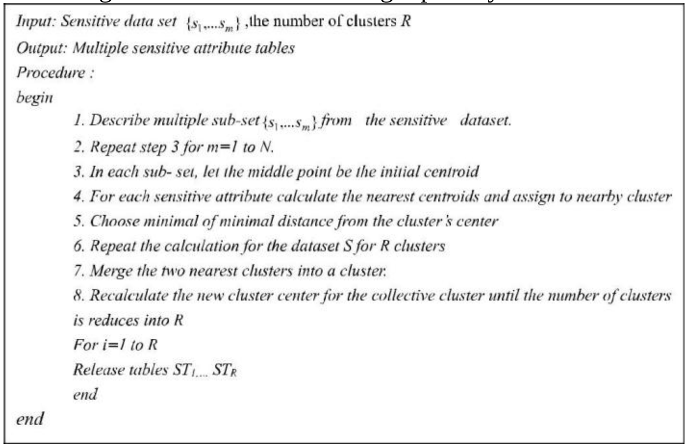
V. PROPOSED WORK

5.1. Anatomizing the dataset D

In order to defeat the defects of generalization, an inventive anatomization technique to attain privacy is presented which yields precise QI values. This phase dissociates the QI and sensitive attributes in micro table T and produces two tables referred to as, QIT and sensitive table. The QI attributes and sensitive attribute have no overlap as the sensitive attributes generally is not seen on publicly available datasets.

5.2. Enhanced slicing algorithm

The slicing algorithm achieves preservation of privacy through horizontal and vertical partitioning. As this work focuses on multiple sensitive attribute, sensitive attribute that are related are grouped together based on their correlation. At that juncture sensitive attribute are sufficiently clustered and results in different tables of sensitive attribute making use of advanced clustering algorithm. And in the subsequent phase tuples are partitioned horizontally by means of MFA and *l*-diversity is checked in for each sensitive tuple. Every sensitive table inserts the correlated attributes along with its group membership within a new column group ID. The partitioning technique removes the dimensionality of the data that ensures this work to be able to deal with any number of sensitive attributes. Finally, the sensitive attribute in each group is shuffled and thereafter linked with a common group id, in such a manner that the sensitive value corresponding to an individual can be found by an intruder with the probability of at the most 1/*l*. A larger *l* leads to a much stronger privacy.



```
Input: Sensitive data set {$s_1,...s_m$} ,the number of clusters R
Output: Multiple sensitive attribute tables
Procedure :
begin
        1. Describe multiple sub-set {$s_1,...s_m$} from  the sensitive  dataset.
        2. Repeat step 3 for m=1 to N.
        3. In each sub- set, let the middle point be the initial centroid
        4. For each sensitive attribute calculate the nearest centroids and assign to nearby cluster
        5. Choose minimal of minimal distance from the cluster's center
        6. Repeat the calculation for the dataset S for R clusters
        7. Merge the two nearest clusters into a cluster.
        8. Recalculate the new cluster center for the collective cluster until the number of clusters
        is reduces into R
        For i=1 to R
        Release tables $ST_{1....}$ $ST_R$
        end
end
```

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 11, November 2017

In a similar way related QI attributes are clustered and its dimensionality is minimized. This provides a new QIT containing correlated attributes in one column and uncorrelated attributes across the column with group membership in a new column group ID.

VI. PERFORMANCE ANALYSIS

The experiments were implemented in Java and carried out on a 2.80 GHz Intel Core processor with 450 GB hard disk and 4 GB RAM having Windows XP operating System. The performance of the algorithm is tested over the datasets obtained from the Cleve-land Clinic Foundation Heart disease and Hungarian Institute of Cardiology, which is available at http://archive.ics.uci.edu/ ml/ datasets/ Heart+ Disease. The data set comprises about 76 raw attributes that decide the probability with regard to the type of patient's heart disease (e.g. class 1 or class 2 or class 3 or class 4). In this technical work, Age, sex, social security number, type of chest pain (Cp), blood pressure at rest (Rbp), serum Scestoral (Sc), blood sugar at fasting (Fbs), electrographic at rest (Restecg), maximum heart rate (Thalach), ST depression due to exercise related to rest (Old peak), exercise induced angina (Exang), slope of the peak exercise ST (Slope), number of major vessels (Ca), blood disorder (Thal), the predicted attribute (Class), which are significant for the ML researchers are taken into consideration. The attributes age, sex, and social security number are considered as QI and the other 12 attributes are considered as sensitive attribute. These experiments have been performed to conceal the 12 sensitive attributes that decide over the probability of the type of the patient's heart disease. The work proposed is realized on both the datasets that are mentioned above. The Cleveland Clinic Foundation Heart disease available at UCI machine learning repository has 303 instances. Since six patient records have more than 25 % of missing values, they are discarded from the dataset. After handling the missing values, the Cleveland dataset is reduced to 297. Similarly, the Hungarian Institute of Cardiology available at UCI machine learning repository has 294 instances and 34 patient records are discarded on account of the missing values. Thus the number of patient records which are taken into account by the proposed work is 557. The comparative evaluation is carried out between the system proposed and the existing MSB (Yang et al. 2008), SLOMS (**SL**icing **O**n **M**ultiple **S**ensitive) (Han et al. 2013) approaches and SLOMS approaches with respect to reconstruction error. As it is expected, while the privacy degree sees an increase, the reconstruction error also finds an increase in all the techniques and the utility is minimized. Figure 1 illustrates the result obtained for the dataset when parameter r is altered. The reconstruction error increases due to high dimensionality and thus limits utility. Then, r = 4 is fixed and the p value is changed. Figure 2 indicates that SLAMSA performs better than MSB and SLOMS approaches with respect to reconstruction error.

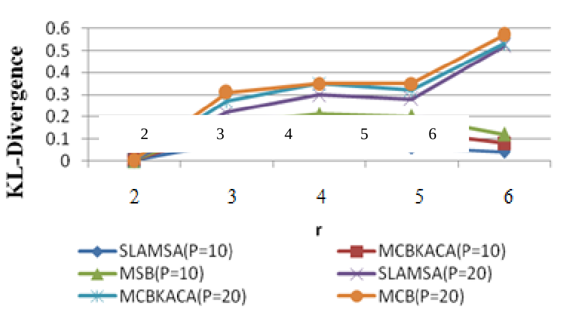


Fig. 1 Reconstruction error versus r (m = 10)

VII. EXECUTION TIME ANALYSIS

The execution time of the proposed work is compared with the prevailing MSB and SLOMS approaches. The execution time of MSB and SLOMS is higher, as it needs to generalize the attributes in each dimension. The anatomization approach eliminates generalization and reduces the execution time by direct release of the QI attributes.

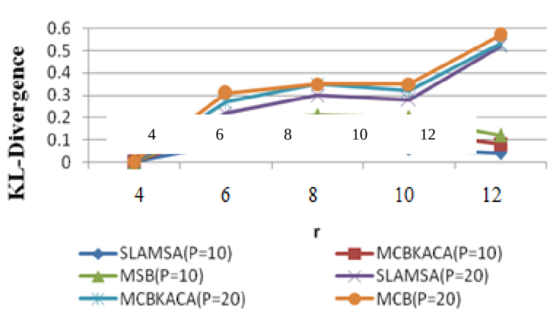


Fig. 2 Reconstruction error versus p (r = 4)

the complexity that is seen in the computation. But the algorithm complexity is high as it does not split the dataset further into several buckets. In order to minimize this problem and for finding the optimal size of buckets, tuple partitioning is formulated by MFA. Additionally, in the proposed work, the single cluster group name is only utilized for referring to several datasets in the samples which also minimizes the complexity of the work. The execution time of the proposed work is measured by means of varying the number of sensitive attributes from 3 to 12. The graphical result is shown in Fig. 3. When the sensitive attribute is one, the execution time is same for all the cases. When the sensitive attribute increases, the proposed techniques is required to process more sensitive attribute and thereafter larger groups are needed to meet the *l*-diversity that again increases the running time.

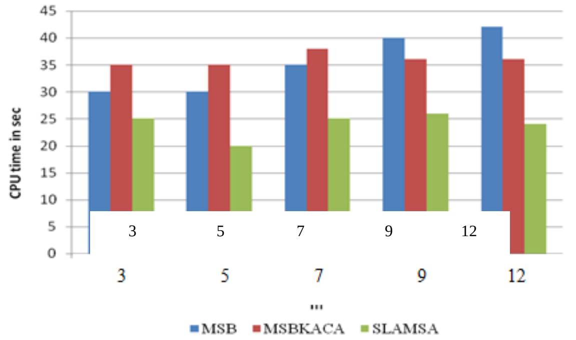


Fig. 3. Execution time with respect to m (l = 3, n = 500)

Experimental results show that SLAMSA keeps up its performance in comparison with other techniques when the sensitive attribute increased. This way, the proposed technique consumes lesser execution time and utility by a significant factor for any number of sensitive attribute in a patients' record.

VIII. CONCLUSION

The important goal of this work is to preserve the privacy of the multiple sensitive attribute and to improve the utility of the health care data. Slicing algorithm helps in preserving correlation and utility and anatomization minimizes the information loss. The advanced clustering algorithms exhibited its efficiency by minimizing the time and complexity. In addition, this work follows the principle of k-anonymity, *l*-diversity. This yields the means for the prevention of privacy threats like membership, identity and attributes disclosure. Also, this method can used to operate for any number of SA in an efficient manner. In future, the slicing algorithm can be applied simultaneously to both QIT and sensitive table to reduce the time further through increased processor speed and memory.

Abbreviations

QIT: quasi identifier table; MFA: metaheuristic firefly algorithm; PPDM: privacy preserving data mining; MSB: multiple sensitive bucketization; MFAMD: Minkowsi distance measure.

REFERENCES

[1] Aggarwal G, Feder T, Kenthapadi K, Motwani R, Panigrahy R, Thomas D, Zhu A. (2005). Anonymizing tables. In: Proceedings of the international conference on database theory (ICDT), vol 3363, Springer, Berlin, pp 246–258.

[2] Clifton C. (2009). Privacy-preserving data mining. In: Liu L, Özsu MT (eds) Encyclopedia of database systems. Springer, US, pp 2147–2150.

[3] El Emam K, Dankar FK. (2008). Protecting privacy using k-anonymity. J Am Med Inform Assoc 15(5):627–637.

[4] Friedman A, Schuster A, Wolff R. (2008). Providing k-anonymity in data mining. Int J Very Large Databases 17:789–804.

[5] Gedik B, Liu L. (2008). Protecting location privacy with personalized k-anonymity: architecture and algorithms. IEEE Trans Mob Comput 7(1):1–18.

[6] Han J, Luo F, Lu J, Peng H. (2013). SLOMS: A privacy preserving data publishing method for multiple sensitive attributes micro data. J Softw 8(12):3096–3104.

[7] http://archive.ics.uci.edu/ml/datasets/Heart+Disease

[8] Iyengar VS. (2002). Transforming data to satisfy privacy constraints. In: Proceedings of the eighth ACM SIGKDD, interna-tional conference on knowledge discovery and data mining, pp 279–288.

[9] Kantarcioglu M, Jin J, Clifton C. (2004). When do data mining results violate privacy? In: Proceedings of the tenth ACM SIGKDD, international conference on knowledge discovery and data mining, 22 August 2004.

[10] LeFevre K, DeWitt DJ, Ramakrishnan R. (2008). Workload-aware anonymization techniques for large-scale datasets. ACM Trans Database Syst 33(3):17:1–17:47.

[11] Li T, Li N. (2006). Optimal k-anonymity with flexible generalization schemes through bottom-up searching. In: Paper pre-sented at sixth IEEE

international conference on data mining workshops, (ICDM workshops) December 2006.
[12] Li T, Li N, Zhang J, Molloy I. (2012). Slicing: a new approach for privacy preserving data publishing. IEEE Trans Knowl Data Eng 24(3):561–574.
[13] Machanavajjhala A, Gehrke J, Kifer D, Venkitasubramaniam M. (2006). L-diversity: privacy beyond k-anonymity. In: Proceed-ings of the 22nd international conference on IEEE, in data engineering (ICDE’06), 24–24.
[14] Meyerson A, Williams R. (2004). On the complexity of optimal k-anonymity. In: Proceedings of the twenty-third ACM SIGMOD-SIGACT-SIGART, symposium on principles of database systems, 14 June 2004.
[15] Pfitzmann A, Hansen M. (2008). Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management: a consolidated proposal for terminology. Version v0 31:1–15.
[16] Shyamala VS, Christopher T. (2015). Managing privacy of sensitive attributes using MFSARNN clustering with optimization technique. J Int Rev Comput Softw 10(9):907–911.
[17] Shyamala VS and Christopher T. (2016). Anatomisation with slicing: a new privacy preservation approach for multiple sensitive attributes. SpringerPlus (2016)5:964.
[18] Tao Y, Chen H, Xiao X, Zhou S, Zhang D. (2009). Angel: enhancing the utility of generalization for privacy preserving publi-cation. IEEE Trans Knowl Data Eng 21(7):1073–1087.
[19] Verykios VS, Bertino E, Fovino IN, Provenza LP, Saygin Y, Theodoridis Y. (2004). State-of-the-art in privacy preserving data mining. ACM SIGMOD Rec 3:50–55.
[20] Xiao X, Tao Y (2006) Anatomy: simple and effective privacy preservation. In: Proceedings of international conference on very large data bases (VLDB), 1 September 2006.
[21] Yang X, Wang YZ, Wang B, Yu G. (2008a). Privacy preserving approaches for multiple sensitive attributes in data publishing. Chin J Comput Chin Ed 31(4):574–587.