

An efficient method for privacy protection in big data analytics using oppositional fruit fly algorithm

Ajmeera Kiran¹, Alwalid Bashier Gism Elseed Ahmed², Mudassir Khan³, J. Chinna Babu⁴,
B. P. Santosh Kumar⁵

¹Department of Computer Science and Engineering, MLR Institute of Technology, Hyderabad, India

²Department of Computer Science, College of Computer Science and Information Technology, University of Bisha, Bisha, Saudi Arabia

³Department of Computer Science, Applied College Tanumah, King Khalid University, Abha, Saudi Arabia

⁴Department of ECE, Annamacharya Institute of Technology and Sciences, Rajampet, India

⁵Department of ECE, YSR Engineering College, YV University, Proddatur, India

Article Info

Article history:

Received Jan 21, 2024

Revised Sep 15, 2024

Accepted Sep 30, 2024

Keywords:

Data privacy
Fruit fly algorithm
Fuzzy C-means
K-anonymization
Oppositional

ABSTRACT

This work employs anonymization techniques to safeguard privacy. Data plays a vital role in corporate decision-making in the current information-centric landscape. Various sectors, like banking and healthcare, gather confidential information on a daily basis. This information is disseminated by multiple sources through numerous methods. Securing sensitive data is of paramount importance for any data mining application. This study safeguarded confidential information using an anonymization technique. Several machine learning methodologies have a deficiency in accuracy. The study seeks to generate superior and more precise results compared to alternative methodologies. For large datasets, numerous solutions exhibit increased time complexity and memory use. For huge datasets, numerous solutions require more time and memory. The enhanced fuzzy C-means (FCM) algorithm surpasses existing approaches in terms of both accuracy and information preservation. This study provides a comprehensive analysis of data anonymization utilizing the oppositional fruit fly approach, a technique that enhances privacy. The clustering method being presented utilizes an enhanced version of the FCM algorithm. The secrecy of the recommended oppositional fruit fly algorithm is effective. The comparison demonstrated that the proposed research enhanced both accuracy and privacy in comparison to two existing methods. The existing strategy outperforms data anonymization-based privacy preservation by 82.17%, while the suggested method surpasses it by 94.17%.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Mudassir Khan

Department of Computer Science, Applied College Tanumah, King Khalid University

Abha, Saudi Arabia

Email: mudassirkhan12@gmail.com

1. INTRODUCTION

Data privacy strategies are categorized into many groups known as data perturbation approaches in multiple research projects. Various data mining techniques are being used to protect the privacy of sensitive personal information that is being kept confidential [1]. Data perturbation, anonymization, random swapping, and cryptographic techniques are examples of procedures that can be used to safeguard an individual's privacy. The value is three. The objective of this study is to propose a privacy protection solution that combines the oppositional fruit fly methodology with data anonymization.

A continuous influx of data is collected from several sources, providing a valuable resource for facilitating decision-making. However, many firms still have challenges in processing data due to the large

volume of information created during their daily operations [2]. Established companies employ diverse data mining techniques to get information about their customers. However, when data is retrieved from databases, there is a potential risk of compromising the confidentiality of sensitive information.

Furthermore, concerns regarding privacy and security in the realm of big data are increasingly evident and pervasive. Legal systems have established laws and regulations to safeguard the confidentiality of persons' information. An illustrious instance of this is the general data protection regulation (GDPR). The GDPR establishes specific criteria for the handling and safeguarding of personal data with the aim of safeguarding individuals' privacy. Data minimization is a fundamental principle of the GDPR, which imposes severe regulations on the storage of data. Consequently, it is advisable to retain personal data for the shortest possible duration necessary to achieve the intended objectives of its processing [3]. Moreover, the GDPR imposes constraints on the utilization of personal data beyond its original collection purpose, highlighting the need of purpose restrictions.

The potential for relaxing these requirements in cases where data has been properly de-identified or anonymized should be duly considered. If there is a data breach, the GDPR advises data controllers and processors to employ de-identification methods to reduce their liability and reporting obligations [4]. By implementing de-identification techniques, companies can strike a harmonious equilibrium between the usefulness of data and the safeguarding of privacy. By following these recommended methods, organizations can guarantee their adherence to privacy regulations while still using the data's value for analysis and decision-making. De-identification is a crucial method that safeguards data privacy and mitigates the dangers associated with the improper use or unauthorized access of personal information [5].

The collection and analysis of data, including sensitive information, are necessary to fully realize certain advantages of information technology. These benefits can only be realized in this manner. Conversely, this could potentially result in undesirable intrusions of personal privacy. Various techniques for safeguarding privacy have been devised to shield the owner from being revealed, achieving this by modifying the initially gathered data [6]. The objective of these measures was to safeguard against the revelation of classified information. However, if the data is altered beforehand, it may result in the extraction of inaccurate or unattainable information through data mining, hence diminishing the data's utility. Privacy-preserving data mining (PPDM) is an acronym for privacy-preserving data mining, which refers to this particular method. The creation of methodologies that incorporate PPDM aims to balance data privacy and information utility. Considering the composition of the remaining components of this piece. The second portion will provide you with an understanding of the essential principles of data mining that safeguard individuals' privacy, along with the main obstacles and historical background of this field. This section presents a clear and detailed explanation and demonstration of the suggested design, together with the data anonymization approach that has been provided.

The organization of the paper includes the introduction section which is presented in section 1. Related works sections are presented in section 2. Anonymization method for proposed system is presented in section 3. Section 4 represents the implementation and experimental results of the proposed method. Section 5 represents the conclusion section.

2. RELATED WORK

Vaidya *et al.* [7] suggested distributed privacy-preserving techniques for partitioning data sets horizontally and vertically using random data transformations (RDTs). Their study outlines the process of constructing protocols, calculating and emphasizing communication expenses, and evaluating security measures. To enhance efficiency and preserve confidentiality, they employed a technique of generating the structure of their system at random. The RDT protocol offers superior privacy protection due to its ability to efficiently handle large amounts of data and its faster processing speed compared to alternative encryption methods. Mendes and Vilela [8] discovered in their survey that computing devices have improved the process of gathering and analyzing data. Data analysis has provided significant advantages to numerous sectors and groups. Storing and transmitting sensitive data provide significant privacy hazards. Kargupta *et al.* [9] developed an approximate random projection-based technique to enhance data privacy while preserving statistical characteristics. The researchers demonstrated the efficacy of the approach in several privacy-preserving data mining applications. An efficient homomorphic encryption and secure comparison mechanism was developed to safeguard personal data. Their research detailed a way for mining frequent item association rules using cloud-based technology. The study by Li *et al.* [10] primarily addresses the integration of disturbances from various participants while maintaining confidentiality in multiparty collaborative mining. This document provides a description of three perturbation-unification processes and their associated costs. Technology enhances scalability, flexibility in data delivery, and privacy by adjusting to spatial constraints. Reducing the number of collaborators may enhance satisfaction but can lead to higher communication expenses. Mehmood *et al.* [11] demonstrated the advantages of utilizing extensive datasets

and the many measures that may be implemented to protect personal data from potential attacks. Jain *et al.* [12] demonstrated the advantages of using large datasets in a complicated structural context to discover intricate connections with different levels of difficulty while maintaining anonymity. The authors utilized T-Closeness, anonymity, and other privacy techniques. Bayan Alabdullah, who is 23 years old, asserts the importance of tradition. Current techniques designed to safeguard small-scale, organized, and uniform data may not be effective in safeguarding non-uniform, disrupted, and exceedingly large quantities of unorganized data. Zhang *et al.* [13] cloud computing can be advantageous for scalable information technology businesses that deal with extremely large datasets. Additionally, the exhibition featured demonstrations of business and healthcare applications, as well as data processing techniques. This research addresses the issues of scalability and local recording in a significant data context with the proposed technique. Sadhwani, *et al.* [14] have stated that the advancements in big data have raised worries regarding privacy and security. In addition, she discusses the difficulties associated with storing large amounts of data. This study investigates various anonymization models designed to safeguard sensitive personal data.

3. ANONYMIZATION METHOD FOR PROPOSED SYSTEM

Big data analytics is the process of storing and retrieving enormous amounts of data from many sources. Information that is sensitive may not be protected from attackers by security [15], [16]. There is a risk of critical data being compromised by using these antiquated methods. Because the data is enormous, our strategy improves the privacy of significant database information [17], to prevent the loss of data during compressing. In order to facilitate scalability, the incoming dataset is grouped. There is a possibility that human organization will lead to data disparities. A modified version of fuzzy C-means (FCM) clustering was utilized. It is not possible to fit all of the data into two or more FCM subgroups. The processing of concurrent data can be sped up by using a mapper to perform clustered data mapping. The data types are transformed via convolution. An approach known as the antagonistic fruit fly method optimizes the classification of artificial neural networks in order to test the correctness of the convolution process [18]. During training, the initial input data is utilized. Repeated convolution is performed in the event that the classification result is lower than the threshold. If the output is greater than a certain threshold, it is sent to a reducer. I then archived it. The semantic configuration of the approach that was suggested is as follows. The k-anonymization architecture that has been suggested is seen in Figure 1. The operation of the map reduction framework is depicted in Figure 2.

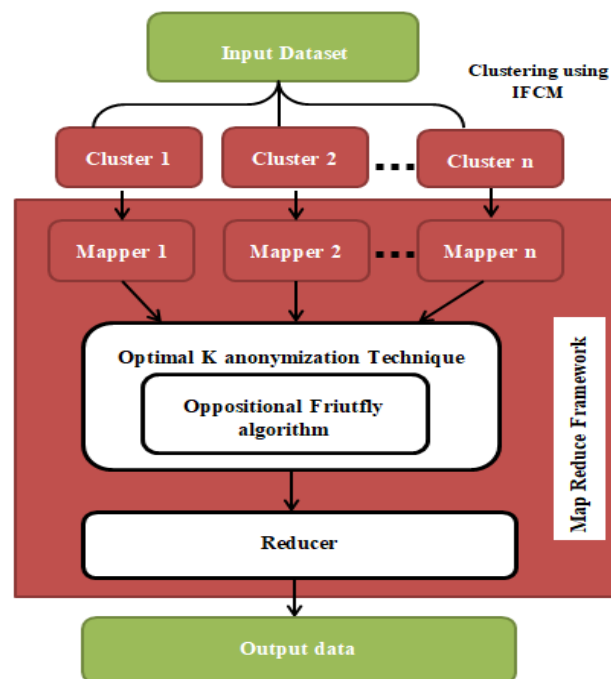


Figure 1. Architecture of proposed k-anonymization approach

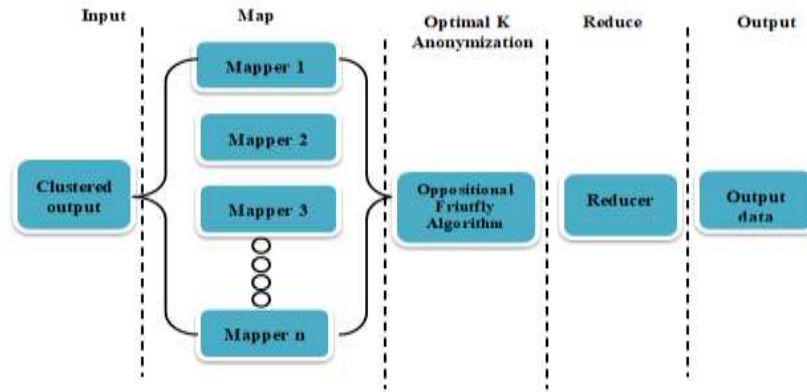


Figure 2. Map-reduce architecture for proposed k-anonymization approach

3.1. Oppositional fruit fly technique

In order to simulate fruit fly foraging, the oppositional fruit fly algorithm (OFA) was developed. The one-of-a-kind algorithm for fruit flies optimizes on a global scale. A fruit fly swarm food foraging investigation was the first step in this process. Hunting for food is the first thing that a fruit fly that possesses both vision and osphresis does. can detect a variety of odors and to fly in the direction of the source of food. Next, it is close to food and has the ability to locate it or travel there using its limited vision. Because the fruit fly algorithm (FA) is based on foraging, optimal solutions are those that suggest food and search for it repeatedly. OFA is an enhanced algorithm that involves fruit flies.

3.2. Enhanced fuzzy C-means method

The objective function is decreased by FCM. Utilizing kernel function for the modification of the FCM algorithm. Due to the fact that the majority of FCM approaches require the cluster count to be pre-specified, they are unable to accommodate even minor cluster modifications. Distances measured using the Euclidean method can sometimes be asymmetrical. The kernel function is responsible for turning the data into a high-dimensional feature space, which in turn improves the performance of the FCM [19]. Unconsciously mapping prototypes to kernel space is an integral part of kernel-based FCM clustering. For the purpose of this article, the Gaussian kernel function ought to be utilized. Among the most common kernel functions are the Gaussian kernel and general kernel functions [20]–[22]. In (1) is used to determine the IFCM mentality objective function, (2) is used to locate centroids, and (3) is used to compute fuzzy membership.

Step 1: the suggested "IFCM's" objective function is shown in expression (1).

$$J(I, C) = \sum_{i=1}^n \sum_{j=1}^c I_{ij}^m (1 - K(D_i, C_j)) \tag{1}$$

Step 2: compute centre C_j :

$$C_j = \frac{\sum_{i=1}^n I_{ij}^m K_G(D_i, C_j) D_i}{\sum_{i=1}^n I_{ij}^m K_G(D_i, C_j)}; j = 1, 2, \dots, C \tag{2}$$

Step 3: calculate fuzzy membership M_{ij} using:

$$M_{ij} = \frac{(1 - K_G(D_i, C_j))^{-1/(m-1)}}{\sum_{i=1}^n (1 - K_G(D_i, C_j))^{-1/(m-1)}}; j = 1, 2, \dots, C \tag{3}$$

Step 4: repeat steps 2 and 3 as many times as you want. Experiment with the Gaussian kernel function to classify the minimization needs with,

$$K_G(D_i, C_j) = \exp(-\|D_i - C_j\|^2 / \sigma^2)$$

where σ is the free parameter. The Euclidean distance is exchanged for a variety of shapes and forms. The preferred way to organize data is as stated above. And then the resultant output is fed to the map reduce framework.

Fan *et al.* [23] explore the enhanced capabilities of the fruit fly optimization algorithm through a boosted hunting mechanism. The paper discusses its application to various real-world problems, highlighting improvements in performance and efficiency. Eckroth [24] discusses the design and implementation of a course focused on big data analytics. The article outlines the curriculum, pedagogical strategies, and the use of various tools and technologies to equip students with the skills necessary for analyzing large datasets. It emphasizes the importance of practical experience and collaborative projects in enhancing learning outcomes. Hewage *et al.* [25] state this systematic literature review examines various privacy-preserving data mining techniques, particularly focusing on their implications for data mining accuracy. The authors, analyze existing methods and their effectiveness in maintaining privacy while ensuring the reliability of data mining outcomes.

4. IMPLEMENTATION AND EXPERIMENTAL RESULTS

In the following part, a description of the experimental data that were obtained for the K-anonymization strategy that was proposed is provided. Our K-anonymization solution, which we have proposed, makes use of a map reduction architecture in order to achieve the aforementioned goal of offering excellent privacy protection for huge data sets. The solution that is being provided makes use of Java, and the Hadoop map is utilized in order to reduce the amount of framework that is required. The IFCM model achieved a success rate of 93.69% when applied to data sizes of 1500 MB, and it achieved a success rate of 94.17% when applied to data sizes of 2000 MB. Because of this, it has been seen that the proposed approach has achieved a high level of accuracy in contrast to the methods that are now being utilized. This is a consequence of the fact that this has occurred. The findings of an accuracy comparison between the proposed IFCM and other approaches are presented in Table 1. The table also includes the outcomes of the comparison.

Figure 3 demonstrates that the improved FCM technique that was proposed was able to achieve an accuracy of 92.88 percent for data that was at least one thousand megabytes in size. This was accomplished by using the technique. An accuracy of 93.69 percent was achieved for data sizes of 1500 megabytes, while an accuracy of 94.17 percent was achieved for data sizes of 2000 megabytes. As shown in Table 1, the correctness of the data that is 2500 megabytes in size is 93.12 percent. This is the conclusion that can be drawn from the data. The conclusion that can be drawn from this is that it has been observed that the proposed method has reached a high level of accuracy in contrast to the methodologies that are currently being utilized. Table 2 presents the results of a comparison between the error value for the proposed IFCM and the error values for existing approaches.

Table 1. Comparison of accuracy for proposed IFCM with other techniques

Data	Accuracy				
	C4.5	RF	K-means	FCM	Improved FCM
1000 MB	74.24	76.28	80.64	90.41	92.88
1500 MB	78.21	80.34	81.34	83.34	92.69
2000 MB	80.51	84.21	85.95	86.35	94.17
2500 MB	79.28	83.45	86.58	91.12	93.12

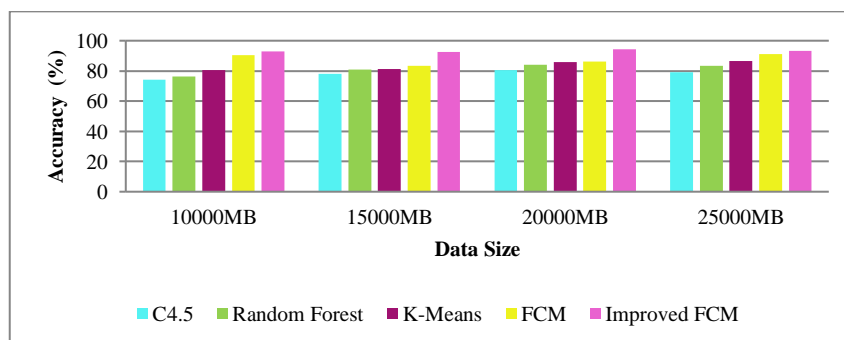


Figure 3. Accuracy comparison of proposed approach with other techniques

Table 2. Evaluation of error value of IFCM with existing methods

Data	Error value comparison				
	C4.5	RF	K-means	FCM	Improved FCM
1000 MB	3.3898	4.0563	4.5968	4.0568	5.3168
1500 MB	4.0545	4.8963	5.0651	5.4562	6.4055
2000 MB	4.8263	4.9856	5.2548	6.1579	7.2263
2500 MB	3.8594	4.0284	4.2563	5.6974	6.4594

Figure 4 provides a visual representation of the comparison of error values about two different groups that are distinct from one another. The graph demonstrates that the solution that was suggested led to an improvement in FCM and ended up reaching an error value of 5.3168 when it was applied to a data size of 1000 MB. For data files that were 1500 MB and 2000 MB in size, respectively, error levels of 6.4055 and 7.2263 were found to be present. A data collection that is 2500 megabytes in size has an error value of 6.4594. This is the last but not the least piece of information. It has been observed that the proposed method has attained a value of error that is minor in contrast to the approaches that are currently in place. This is a consequence of the fact that this has occurred.

As demonstrated in Figure 5, the recommended strategy enhanced FCM by gaining a total of 13615478 bits of memory. This was accomplished by utilizing the proposed method. To achieve this, a data capacity of one thousand megabytes was utilized. The amount of random-access memory (RAM) that is required for data sizes of 1500 MB and 2000 MB, respectively, is 11236598 bits and 10698325 bits, respectively. The quantity of memory space that is required for data that is 2500 megabytes in size is 9632541 bits of memory, as stated in the conclusion. Therefore, when compared to the methods that are now being utilized, it is noted that the suggested technique utilizes less memory space as a natural consequence. This is the case since the proposed method is more efficient. The findings of a comparison between the memory consumption of the suggested IFCM and those of other possible techniques are presented in Table 3, which illustrates the outcomes of the comparison

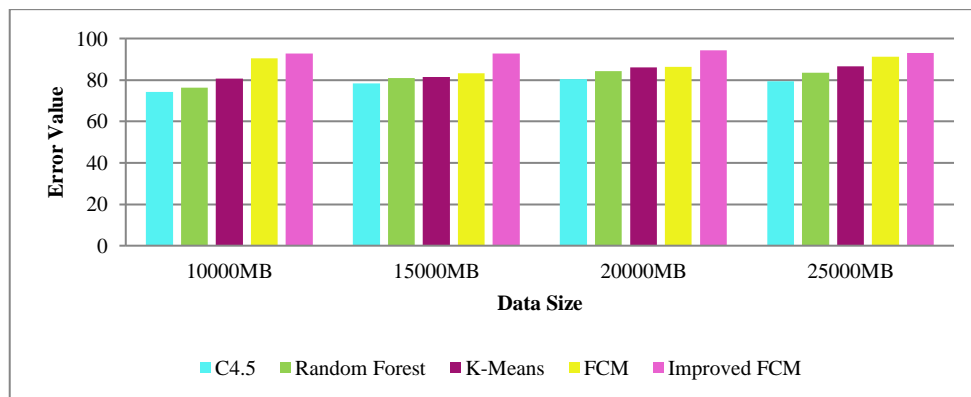


Figure 4. Evaluation of error value of IFCM with existing methods

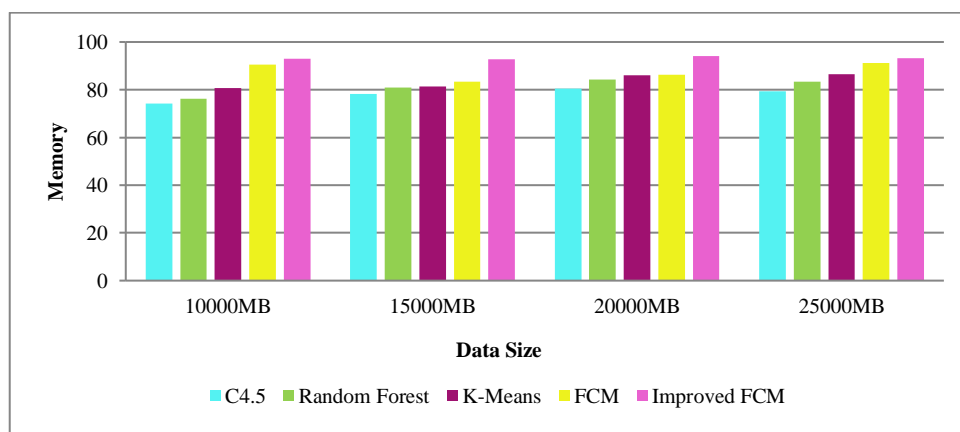


Figure 5. Evaluation of memory value of IFCM with existing methods

Table 3. Evaluation of memory value of IFCM with existing methods

Data	Memory (bits)				
	C4.5	RF	K-means	FCM	Improved FCM
1000 MB	18325421	17654214	16542132	15423657	13615478
1500 MB	16236528	15365232	14657498	12986541	11236598
2000 MB	15326541	14235412	13965874	11362498	10698325
2500 MB	14658974	13544659	11326541	10965874	9632541

Figure 6 provides a visual representation of the comparison of error values that may be found in the chart. The following graphic illustrates how the proposed solution enhanced FCM to the point where it reached 3,269,325 milliseconds in time for a data capacity of 1000 gigabytes. This improvement may be seen in the diagram provided below. It takes 3,685,475 milliseconds and 3,063,598 milliseconds, respectively, to complete the task when the data size is 1500 megabytes and 2000 megabytes. In conclusion, the amount of time required to process 2500 gigabytes of data is 3,959,421 minutes for the entire procedure. On account of this, the proposed strategy demands a shorter amount of time in contrast to the methodologies that are now being utilized. Table 4 is the name of the table that provides a comparison of the amount of time that is required by the proposed IFCM in comparison to that of other techniques.

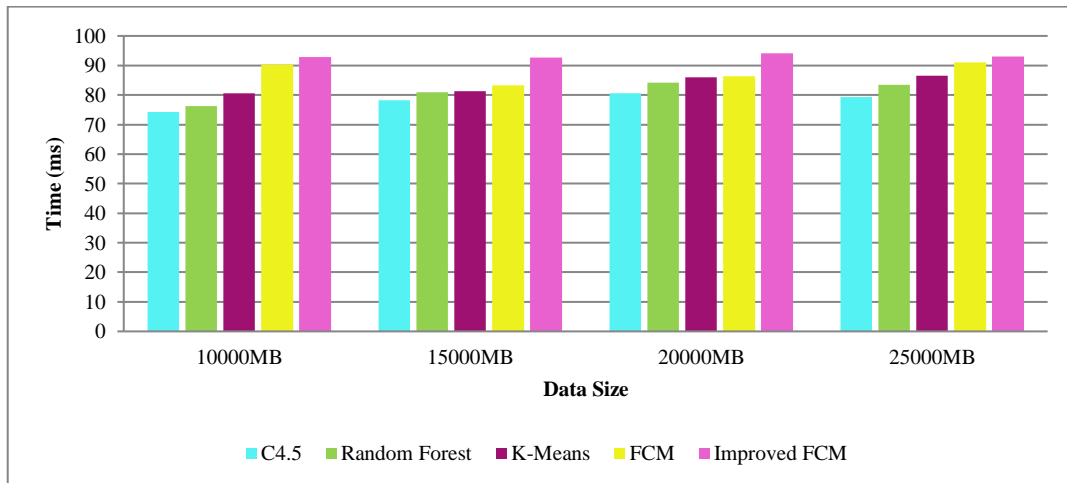


Figure 6. Evaluation of time consumed of IFCM with existing methods

Table 4. Evaluation of time consumed of IFCM with existing method

Data	Time (ms)				
	C4.5	RF	K-means	FCM	Improved IFCM
1000 MB	4363325	4136958	3954875	3563242	3269325
1500 MB	4238763	4021548	3869857	3687541	3685475
2000 MB	4013696	3963599	3636584	3323654	3063598
2500 MB	5295587	4959745	4469852	4186541	3959421

The information loss that occurs when utilizing the proposed IFCM is compared to the information loss that occurs when using alternative techniques, and the results are presented in Table 5. The strategy that was proposed indicates a lower amount of information loss when compared to other methods, such as PKA (particle K-anonymization), K-anonymization, SKA (scalable K-anonymization), and others. This is displayed in Figure 7, which shows the comparison. The graph illustrates that the quantity of information that is lost is depicted as being significantly less when compared to the procedures that are currently in place.

Table 5. Information loss comparison for proposed IFCM with existing classifier techniques

Metrics	PKA [18]	KA [18]	SKA [19]	Improved FCM
Information Loss	45	50	45	32

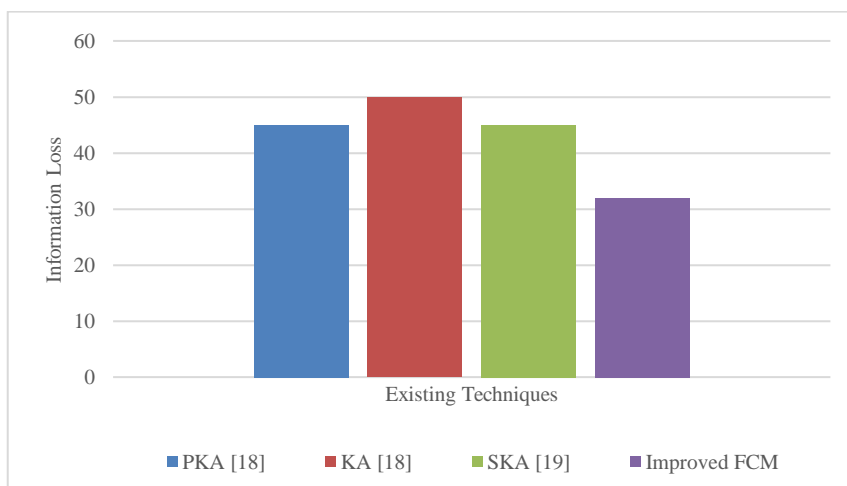


Figure 7. Information loss comparison for proposed IFCM with existing classifier techniques

5. CONCLUSION

This research makes use of data anonymization and the fruit fly approach, which is the opposite of the fruit fly method. The proposed method is evaluated in terms of memory, execution time, accuracy, and error value for data storage capacities of 10,000 MB, 15,000 MB, 20,000 MB, and 25,000 MB respectively. The process of anonymizing data is one of the technologies that is both the quickest and the most secure for protecting privacy. Estimating the anonymization threshold k is accomplished through the utilization of the oppositional fruit fly algorithm by the privacy-preserving technique employed. The methods used in machine learning have not been successful in improving data accuracy at the expense of greater information loss. In many systems, large datasets both take more time and require more memory to process. With the updated FCM, it is possible to obtain greater accuracy and less information loss than with the techniques that are currently in use. There is a difference of 82.17% between the performance of the present approach and that of the suggested method when it comes to data anonymization. The subsequent investigation has the potential to enhance privacy by utilizing the mechanisms that are already in place. It is possible to enhance security by expanding the use of convex optimization. Our proposed IFCM methods are compatible with a limited number of datasets; however, we are working to expand our coverage to include additional types of data. The data is likewise lost using our procedure. Our objective is to reduce the amount of data that is lost in the future.

ACKNOWLEDGMENTS

The authors are thankful to the Deanship of Graduate Studies and Scientific Research at University of Bisha for supporting this work through the Fast-Track Research Support Program.





REFERENCES

- [1] Zhao, C. Zhang, and S. Guan, "A data lake-based security transmission and storage scheme for streaming big data," *Cluster Computing*, vol. 27, no. 4, pp. 4741–4755, 2024, doi: 10.1007/s10586-023-04201-9.
- [2] A. Kiran and D. Vasumathi, "A comprehensive survey on privacy preservation algorithms in data mining," *2017 IEEE International Conference on Computational Intelligence and Computing Research (ICIC)*, Coimbatore, India, 2017, pp. 1-7, doi: 10.1109/ICIC.2017.8524294.
- [3] S. Uthayasankar, M. M. Kamal, Z. Irani, and V. Weerakkody, "Critical analysis of big data challenges and analytical methods," *Journal of Business Research*, vol.70, pp. 263-286, 2017, doi: 10.1016/j.jbusres.2016.08.001.
- [4] A. Kiran and D. Vasumathi, "Optimal privacy preserving technique over big data analytics using oppositional fruit fly algorithm," *Recent Advances in Computer Science and Communications*, vol. 13, no. 2, pp: 283-295, 2020, doi: 10.2174/2213275911666181119113913.
- [5] A. Oussous, F.-Z. Benjelloun, A. A. Lahcen, and S. Belfkih, "Big data technologies: A Survey," *Journal of King Saud University-Computer and Information Sciences*, vol. 30, no. 4, pp. 431-448, 2018, doi: 10.1016/j.jksuci.2017.06.001.
- [6] A. Kiran and D. Vasumathi, "Data mining: min–max normalization-based data perturbation technique for privacy preservation," *Proc Fourth International Conference on Computational Intelligence and Informatics*, pp.723-734, 2020, doi: 10.1007/978-981-15-1480-7_66.
- [7] J. Vaidya, Y. M. Zhu, and C. W. Clifton, "Privacy and data mining," *Privacy Preserving Data Mining. Advances in Information Security*, vol. 19, Springer, Boston, ISBN: 978-0-387-29489-6, 2006.
- [8] R. Mendes and J. P. Vilela, "Privacy-preserving data mining: methods, metrics, and applications," *IEEE Access*, vol. 5, pp. 10562-10582, 2017, doi: 10.1109/ACCESS.2017.2706947.





- [9] H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar, "Random-data perturbation techniques and privacy-preserving data mining", *Knowledge and Information System*, pp. 387-414, 2005, doi: 10.1007/s10115-004-0173-6.
- [10] Y. Li, M. Chen, Q. Li and W. Zhang, "Enabling multilevel trust in privacy preserving data mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 9, pp. 1598-1612, 2012, doi: 10.1109/TKDE.2011.124
- [11] A. Mehmood, I. Natgunanathan, Y. Xiang, G. Hua, and S. Guo, "Protection of Big Data Privacy," *IEEE Access*, vol. 4, pp. 1821-1834, 2016, doi: 10.1109/ACCESS.2016.2558446.
- [12] P. Jain, M. Gyanchandani, and N. Khare, "Big data privacy: a technological perspective and review," *Journal of Big Data*, vol. 3, no. 1, pp.1-25, 2016, doi: 10.1186/s40537-016-0059-y.
- [13] X. Zhang, W. Dou, J. Pei, S. Nepal, C. Yang, and C. Liu, "Proximity-aware local-recoding anonymization with MapReduce for scalable big data privacy preservation in cloud," *IEEE Transactions on Computers*, vol. 64, no. 8, pp. 2293-2307, 1 Aug. 2015, doi: 10.1109/TC.2014.2360516
- [14] D. Sadhwani, S. Silakari, and U. Chourasia, "Preserving privacy during big data publishing using K-anonymity model-a survey," *International Journal of Advanced Research in Computer Science*, vol. 8, no. 5, 2017, doi: 10.26483/ijarcs.v8i5.3426.
- [15] N. Kshetri, "Big data's impact on privacy, security and consumer welfare," *Telecommunications Policy*, vol. 38, no. 11, pp. 1134-1145, 2014, doi: 10.1016/j.telpol.2014.10.002.
- [16] J. Hu and A. V. Vasilakos, "Energy big data analytics and security: challenges and opportunities," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2423-2436, Sept. 2016, doi: 10.1109/TSG.2016.2563461.
- [17] L. Chen, C. L. P. Chen and M. Lu, "A multiple-Kernel fuzzy C-means algorithm for image segmentation," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 5, pp. 1263-1274, Oct. 2011, doi: 10.1109/TSMCB.2011.2124455.
- [18] S. N. Bushra and A. Chandrasekar, "Privacy preservation on big data using PK-anonymization," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 3, no. 11, pp. 11937-11942, 2015.
- [19] M. B. Brijesh and U. P. Rao, "Privacy-preserving big data publishing: a scalable k-anonymization approach using Map-Reduce," *IET Software*, vol. 11, no. 5, pp. 271-276, 2017, doi: 10.1049/iet-sen.2016.
- [20] A. Abbasi and B. Mohammadi, "A clustering-based anonymization approach for privacy-preserving in the healthcare cloud," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 1, 2022, doi: 10.1002/cpe.6487.
- [21] N. Rogovschi, Y. Bennani, and S. Zouinina, "Data anonymization through multi-modular clustering," *Recent Advancements Multi-View Data Analytics*. Cham, Switzerland: Springer, 2022, pp. 159-176.
- [22] S. Srijayanthi and T. Sethukarasi, "Design of privacy preserving model based on clustering involved anonymization along with feature selection," *Computer Security*, vol. 126, Mar. 2023.
- [23] Y. Fan *et al.*, "Boosted hunting-based fruit fly optimization and advances in real-world problems," *Expert Systems with Applications*, vol. 159, 2020, doi: 10.1016/j.eswa.2020.113502.
- [24] J. Eckroth, "A course on big data analytics," *Journal of Parallel and Distributed Computing*, vol. 118, pp. 166-176, 2018, doi: 10.1016/j.jpdc.2018.02.019.
- [25] U. H. W. Hewage, R. Sinha, and M. A. Naeem, "Privacy-preserving data (stream) mining techniques and their impact on data mining accuracy: a systematic literature review," *Artificial Intelligence Review*, vol. 56, pp. 1-38, doi: 10.1007/s10462-023-10425-3.

BIOGRAPHIES OF AUTHORS







Ajmeera Kiran     working as assistant professor, Computer Science and Engineering Department at MLR Institute of Technology, Hyderabad, Telangana, India. He received the B.E degree in Computer Science and Engineering from Vasavi College of Engineering affiliated to Osmania University, Hyderabad in 2012. In 2014, received the M.Tech. degree in Computer Science and Engineering from Jawaharlal Nehru Technological University, Hyderabad. In 2021, received the Ph.D. degree in Faculty of Computer Science & Engineering from JNTUH, Hyderabad, Telangana, India. He has 3-year experience in teaching and 6.5 years in research, he published around 42 publications in international/national journals and conferences. 8 patents in data mining and big data analytics. He also authored 6 books/books chapters by various international publishers. He was also a convener for various FDP's, coordinator for student workshop's, Hackathon's. He is a life member of professional bodies like Indian Society for Technical Education (LMISTE), Member in IEEE. Research interest includes information security, network security, big data analytics and machine learning, artificial intelligence, data mining, and internet of things. He can be contacted at email: kiranphd.jntuh@gmail.com.







Alwalid Bashier Gism Elseed Ahmed     his bachelor Bachelor's degree from Al-Neelain University of Science and Technology in 2002. He received Master's degree from Al-Neelain University-College of Computer Science and Information Technology in 2008 and Ph.D. degree from Al-Neelain University-College of Graduate Studies in 2015. His major research area includes E-government's applications, artificial intelligence, and big data. He can be contacted at email: alwldbasheer@ub.edu.sa.







Mudassir Khan     is currently working as assistant professor and former Head in the Department of Computer Science at College of Science and Arts Tanumah, King Khalid University Abha, Saudi Arabia. He has completed Ph.D. in Computer Science from Noida International University (NIU), India. He has completed his Graduation from Aligarh Muslim University and master's from Gautam Budh Technical University, India. He has more than 13+ years of Teaching Experience at the King Khalid University of Saudi Arabia. He has published more than 50+ papers in International Journals (SCIE, ESCI, Scopus, UGC Care), conferences IEEE and Springer Nature) and 4 patents. He is the Member of various technical/ professional societies such as IEEE, UACEE, Internet Society, IAENG and CSTA. His research interest includes big data, IoT, deep learning, machine learning, e-learning, fuzzy logic, image processing, cyber security, and cloud computing. He can be contacted at email: mudassirkhan12@gmail.com.



J. Chinna Babu     has received his B. Tech. degree from JNTU, Hyderabad. M.Tech. degree in VLSI System Design and Ph.D. in VLSI signal processing degrees from JNTUA, Ananthapuramu. Currently, he is working as Associate Professor in the Department of Electronics and Communication Engineering at Annamacharya Institute of Technology and Sciences-Autonomous, Rajampet, Kadapa, A.P, and India. He has published number of research papers in various international journals and various national and international conferences. He is a review member of various international journals. He was joined at AITS, Rajampet as an Assistant Professor in the Department of Electronics and Communication Engineering in the year 2008. He is having 15 years of Experience in Teaching UG and PG courses. He is a life Member in ISTE and Member in IACSIT, MIREN and CSTA. His current research interests are VLSI signal processing, microprocessor, embedded systems and communication, IoT, machine learning, and signal processing. He can be contacted at email: jchinnababu@gmail.com.



B. P. Santosh Kumar     is currently working as Associate Professor and Head, Department of Electronics and Communication Engineering at YSR Engineering College of Yogi Vemana University, Proddatur, India. He has 18 years of teaching experience and under his guidance 04 research scholars are working in the image processing fields. He received B.Tech. Degree from JNTUH in 2001, M.Tech. Degree from Kerala University, Thiruvananthapuram in 2006 and Ph.D. degree from Yogi Vemana University, Kadapa in 2020. He has published 33 papers in international journals, 10 papers in international conferences, 07 books in international publication, 03 book chapters and 04 patents. He is a lifetime member of various professional bodies like ISTE, MIE. His current research interests include signal processing, image processing, biomedical signal, and image processing. He can be contacted at email: santoshyvuce@gmail.com.