American Journal of Bioscience and Bioengineering
2023; 11(5): 79-91
http://www.sciencepublishinggroup.com/j/bio
doi: 10.11648/j.bio.20231105.14
ISSN: 2328-5885 (Print); ISSN: 2328-5893 (Online)

Enhanced Privacy Preservation Technique for the Multi Institutional Clinical Data Using Hybrid Feline-Storm Algorithm

Sagarkumar Patel[1, *], **Rachna Patel**[2, *]

[1]Department of Biometrics, LabCorp Drug Development Inc, New Jersey, USA

[2]Department of Biometrics, Catalyst Clinical Research Llc, North Carolina, USA

Email address:

Sagarkumar.patel1@fortrea.com (Sagarkumar Patel), Rachna.patel@catalystcr.com (Rachna Patel)

[*]Corresponding author

To cite this article:

Sagarkumar Patel, Rachna Patel. Enhanced Privacy Preservation Technique for the Multi Institutional Clinical Data Using Hybrid Feline-Storm Algorithm. *American Journal of Bioscience and Bioengineering*. Vol. 11, No. 5, 2023, pp. 79-91. doi: 10.11648/j.bio.20231105.14

Received: September 14, 2023; **Accepted**: September 26, 2023; **Published**: October 9, 2023

Abstract: Today's medical research is seen to be highly dependent on data exchange; unfortunately, despite its benefits, it frequently encounters problems, particularly issues with data privacy. As a result, several methods and infrastructures have been created to ensure that patients and research participants maintain their anonymity when data is exchanged. However, privacy protection often has a cost, such as limitations on the types of studies that may be done on shared data. The lack of a systematization that would make the trade-offs made by various techniques obvious is what needs to be addressed. In this research, develop the Feline-Storm Based Privacy Preservation Technique for multi-institutional clinical data. Data mining provides many advantages in various domains, particularly in medicine. The data about the disease is ensured to the experts, who can determine the effects, availability, and nature. The private information of the persons should not be disclosed to the expert groups, which ensures the confidentiality of the confidential information. Hence, to ensure the privacy of the people's electronic health records (EHR), this research utilizes the C-mixture and three privacy restraints that strengthen the privacy measures. Furthermore, the Hybrid Feline-storm algorithm, which emphasizes exploitation or the exploration phase at any instance, avoiding the local optima and the premature convergence to ensure the optimized privacy preserved of the data. This research also establishes security strategies such as K-anonymity, T-closeness, and L-diversity to attain complete data privacy. Further, the Feline-storm optimization is developed to minimize information loss. The information loss, class average size, and fitness measure achieved by the proposed methodology are 0.85, 0.38, and 4.7457, respectively.

Keywords: Hybrid Feline-Storm Algorithm, Data Anonymization, C-mixture, Clinical Trial and Pharma Industry

1. Introduction

Artificial intelligence (AI) systems that can help physicians in several situations, such as the early diagnosis of tumors in medical imaging, have lately shown hopeful outcomes due to the rapid advancement of AI and machine learning (ML) in biomedical data analysis. As seen by the increasing number of regulatory approvals [1] and patent applications [2], these systems are developing past the proof-of-concept stage. They are anticipated to find broad use in the upcoming years. The demand for extensive and diverse datasets for training the ML models is a feature of high-performance AI systems, and this requirement is frequently met by voluntary data sharing on the part of data owners and the aggregation of datasets across institutions. Patients' data is frequently anonymized or pseudonymized at the institution where it was collected, then sent to and maintained at the location of analysis and model training (a process known as centralized data sharing) [3-6]. Anonymization, however, has shown to be ineffective at defending against re-identification attempts [7, 8]. In context, it is essential from a moral and legal perspective to gather, aggregate, and transmit patient data on a broad scale [9]. Additionally, controlling personal health data's transmission, storage, and use is a core patient right. This control is virtually eliminated by centralized data sharing, which results in a loss of sovereignty.

Furthermore, once transferred, anonymized data cannot be easily updated or changed in the past, for instance, by adding new clinical data that becomes available. Despite these issues, the growing desire for data-driven solutions is projected to lead to a rise in the collection of health-related data, including through mobile devices and wearable health sensors in addition to clinical records, hospital patient data, and information from medical imaging [10-13]. To reconcile data and safeguard privacy, innovative approaches are therefore needed. While still allowing for valuable inferences from the data or its usage for model construction, secure and privacy-preserving machine learning (PPML) strives to maintain data security, privacy, and confidentiality. Despite the limited local data availability, PPML allows for the creation of advanced models in low-trust contexts [14-18].

IoT denotes a progressive innovation that influences routine life through various applications, assuring a safe, innovative, and more accessible environment [19-20]. IoT facilitates the interconnection between numerous individuals by integrating billions of people, substantial objects, services, intelligent nodes, and digital sensors. The IoT establishes persistent information transmission and mutual interaction between things, consisting of millions of digital sensors [21]. The significant feature of IoT is that all the objects in the environments are interlinked to each other as it transfers the data in any part of the world at any time [22, 23]. Modern innovations such as big data, cloud computing, fog computing, allocated computing, and wireless communication help IoT again to facilitate interaction between intelligent objects [24-26]. The wide range of applications such as transport, smart home, health care sector, and electricity conservation enables IoT devices [27] to generate copious amounts of big data. In the medical field, data related to patients are gathered and stored in EHRs [28, 29]. The EHR records incorporate various data such as admission, discharge, and Transfer (ADT) of the patient along with routine clinical checkups, patient's confidential information, medical and genealogical history, genomic sequence, immunization, and therapeutic communication, patronage information, and other beneficial data regarding the patient [30-32]. Though these big data play a significant role in the healthcare sector, the technological progressions and advantages following the accumulation of enormous amounts of information in the medical care industry are not impervious to many security-related difficulties. Preserving the privacy and the security of confidential data is a significant challenge. Different innovations guarantee the safety and protection of critical medical services information. Yet, these system experiences technical issues such as security threats, decentralized architecture, and low reconstruction capacity [33, 34].

The main aim of the research is to develop the Feline-Storm Based Privacy Preservation Technique for multi-institutional clinical data to ensure the privacy of the people's EHR records; this research utilizes the C-mixture and three privacy restraints that strengthen the privacy measures. Furthermore, the Hybrid Feline-storm algorithm, which emphasizes exploitation or the exploration phase at any instance, avoids the local optima and the premature convergence to ensure the optimized privacy preserved of the data. This research also establishes various security strategies such as K-anonymity, T-closeness, and L-diversity to attain complete data privacy. Further, the Feline-storm optimization is developed to minimize information loss.

Feline-Storm Algorithm:

The proposed calculation is developed to integrate the two-principal behavior of felines and humans' mutual interaction. A feline appears to be sluggish and spends a more significant part of their time resting. Still, their awareness is exceptionally high during their rests, and they are conscious of what transpires around them. Thus, they continually notice the environmental factors brilliantly and intentionally as soon as they see their objective, they begin moving toward it rapidly. On the other hand, humans can sort out the best solution for any problem through mutual interactions.

2. Literature Review

To solve the multi-site fMRI classification issue, implement an approach that protects privacy. Xiaoxiao Li et al. [35] created a federated learning strategy in which a shared local model's shared weights are modified using a randomization mechanism. But acquiring a lot of data at one location is challenging because of big fMRI datasets. The collaborative deep learning system created by Ling Chen Zhao et al. [2] allows users to work together to create a collective deep learning model using the data of all participants without the need for direct data sharing or centralized data storage. Although our model was highly accurate and resilient to unreliable participants, it had difficulty gathering all the participant data upfront and training on the complete dataset.

A Fed GRU algorithm was created by Jiawen Kang and Dusit Niyato [1] for traffic flow prediction to protect privacy. According to this model, the Fed GRU's prediction accuracy was higher than that of more complex deep learning models, which makes it more challenging for the cloud to carry out gradient information aggregation concurrently. A blockchain-enabled safe data-sharing architecture for dispersed multiple parties was created by Yunlong Lu [3], and it achieved excellent efficiency, improved security, and good accuracy. Due to their constrained computation and storage capabilities, IIoT end devices find it difficult to output and maintain structured data. Imran Makhdoom et al. [7] created an inventive framework based on blockchain for privacy-preserving and secure IoT data exchange in a smart city context. This model achieved high latency and low throughput however, this approach looks to have significant communication complexity.

A PriMIA (Privacy-preserving Medical Image Analysis) model was developed by Georgios Kaissis et al. [8], an open-source software framework and encrypted inference on medical imaging data. This model had high communication efficiency, but it also had more computational complexity.

A federated learning model for multi-institutional collaboration was created by Micah J. Sheller et al. [9];. In contrast, this model showed excellent accuracy compared to previous models; sending medical data to a centralized site in this approach presents several legal, privacy, technical, and data-ownership difficulties. Devendra Dhagarra et al.'s [10] goal is to comprehend the connections that predict patients' acceptance of technology in healthcare services. This concept raised significant privacy issues regarding the use of technology in healthcare, but it also encountered numerous computational challenges.

2.1. Deep Learning

Broadly speaking, deep learning, based on artificial neural networks, aims to learn and extract high-level abstractions in data and build a network model to describe accurate relations between inputs and outputs [2]. Common deep learning models are usually constructed by multi-layer networks, where non-linear functions are embedded, so that more complicated underlying features and relations can be learned in different layers [2].

2.2. Challenges

1) Collecting all participant data in advance and training on the complete dataset has several flaws and challenges. High transmission costs are typically associated with such a convoluted data collection method.
2) However, the fewer nodes also make it more challenging to reach consensus, and it is also difficult to attract a large number of candidates to a single location due to the time and expense required for dataset capture and annotation.
3) It provides problems with dataset imbalance, complex image augmentation, and federation-wide hyperparameter tuning functionality that are frequent in medical imaging analytic workflows.
4) A significant challenge in medical imaging is obtaining enough data and data imbalance.

3. Methodology

Data mining provides many advantages in various domains, particularly in medicine. The private information of the persons should not be disclosed to the expert groups, which ensures the confidentiality of the confidential information. Hence, to ensure the privacy of the people's EHR records, this research utilizes the C-mixture and three privacy restraints that strengthen and enhance the privacy measures. Furthermore, the Hybrid Feline-storm algorithm, which emphasizes exploitation or the exploration phase at any instance, avoiding the local optima and the premature convergence to ensure the optimized privacy preserved of the data. A quantitative research design is being used for this current study. The research focuses on various privacy and security issues in recent technical innovations. Despite all the advantages associated with distributed computing applications for medical services, various health sector innovations, safety efforts, and legal issues must be addressed.

Consequently, there is a need to carry out privacy protection techniques to forestall multi-institutional clinical data. The dissertation uses the feline-storm-based Privacy Preservation Technique with other conventional methods such as Genetic algorithm, GWO algorithm, Genetic + GWO algorithm, CSO algorithm, and BSO algorithm. The comparative evaluation using the Cleveland data is illustrated in the section in terms of $GIn_{loss}Class_{avg}$, and Fitconcerning the C-mixture value. The multi-institutional clinical data must be transferred from one institution to the other clinical institution to obtain expert advice. Therefore, the intrusion of unauthorized users will violate the privacy of the patient's data. Hence, many researchers concentrate on developing a dynamic attack-resilient privacy protection technique in multi-institutional data. This research establishes security strategies such as K-anonymity, T-closeness, and L-diversity to attain complete data privacy.

Further, the Feline-storm optimization is developed to minimize information loss. Figure 1 depicts the outline of the privacy-preserved data-distributing methodology. The data from the people saved to such an extent that those records require security and privacy. The term privacy clarifies that the outsider doesn't connect the singular information.

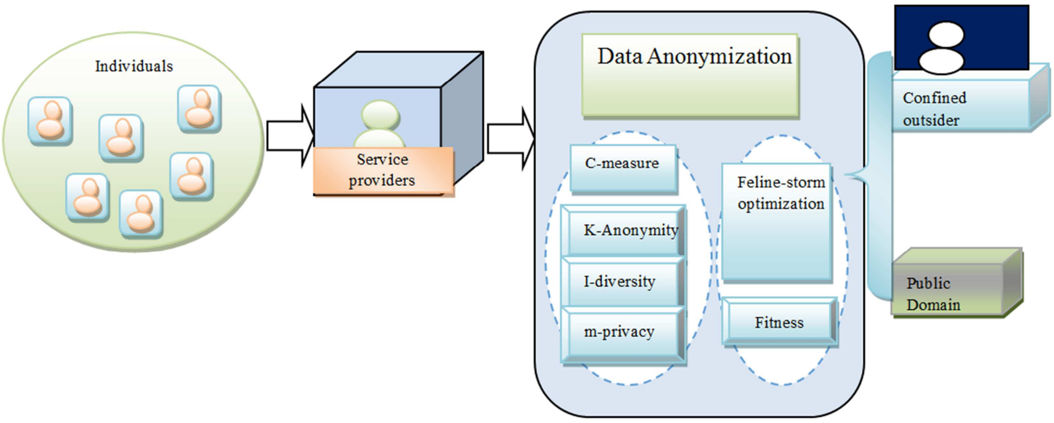


Figure 1. *Architecture of the proposed preserving mode.*

3.1. System Model of IoT and Privacy Concerns

In this research, an IoT network is established through the simulation in MATLAB, in which each node gathers data from a different medical institution. The collected information or the patient's clinical data from various healthcare institutions are now stored in the cloud through the base station. Recently, several researchers have focused on developing an effective privacy protection technique for multi-institutional data. Hence, advanced anonymization techniques are established in this research to enhance the privacy of the multi-institutional clinical data.

3.1.1. Model for Privacy-Preserved Data

This section elucidates the proposed data anonymization model to enhance personal data privacy in multi-institutional health data. The outsiders will obtain data from the service providers' semi-identifiers, names, and sensitive and non-sensitive attributes. The data provided by the service providers are mathematically represented in the following equation:

$$D = \{D_n \in S_n; 1 \leq n \leq N\} \tag{1}$$

D_n denotes the report conferred by n^{th} service provider, S_n is the n^{th} service provider, $n$is the number of providers engrossed in entrusting their EHR to the confined outsiders, and Nsignifies the total number of providers authorizing their reports. The individual EHR is represented in Equation 1. The EHR published by the providers consists of semi-identifiers and attributes, which are illustrated in the following equation:

$$D = \{S, a_n, a_2, a_o, Iq_n, Iq_2, Iq_o, sa_n, sa_2, sa_o\} \tag{2}$$

Where Srefers to the name of the providers, $a$denotes the common attributes sa represents the sensitive attributes, and $Iq$is the semi-identifier The confined outsider distributes the report given by the providers through the alterations made to D and the report to be distributed introduced asD^*. The report D^*Restrains all the assaults that are executed to recognize the individual data. The distributed data is safeguarded so that any assault on the distributed information never reveals the personal data or gives any character to the confidential information. The privacy of the information from exposing the report's contents to the individual data is saved through the accompanying strategies known as k-anonymity, m-privacy, and l-diversity. The security procedures improve by utilizing the speculation interaction that makes the information less explicit.

3.1.2. C-Mixtures for Upgrading the Privacy of Data

For EH report D encompassing the semi-identifiers, Iq the k-anonymity expresses that the report D fulfills k-anonymity regarding the identifier Iq if and only if, for each arrangement in$D[Iq]$, as there are a minimum k number of events in$D[Iq]$. The k-anonymity goes through speculation to generate $k-1$ undefined records and is less instructive. The procedure for making a report indefinitely gets the report from connecting a person. Yet, k-anonymity alone cannot guarantee total security to the information since it exposes background assaults and experiences homogeneity. Therefore, other privacy measures such as m-privacy and l-diversity are also utilized in the study. A combination of these approaches is co-operatively utilized to preserve the confidentiality of the data. The information safeguards privacy uses a boundary C named the C-mixture introduced in the following section. C-mixture estimates the secrecy of the distributed information that alters the three protection measures. Any information that is distributed is required to fulfill the requirements to guarantee data privacy. Consider the EH consists of the semi-identifierIs, and then the report must satisfy the condition of the C-mixture.

Genetic grey wolf optimization and C-mixture for collaborative data publishing

1) There should be at least C% identical records in each gathering.
2) It is essential to have C% distinct qualities for sensitive attributes (Sa) of each gathering,
3) There ought to be C% providers for each gathering. Then, utilizing the C-mixture, the security constraints are formula.

$$[k = \tau_n * C][l = a_n * C][m = N * C] \tag{3}$$

Where τ_n demonstrates the total number of records, C shows the indication of C-mixture, a_n Demonstrates the number of attributes and Nreflects the number of providers.

3.1.3. Hybrid Model- Feline Storm Algorithm for Establishing Privacy in Health Care Data

This section elucidates the hybrid algorithm known as the Feline storm optimization algorithm for establishing healthcare data privacy. Optimization occurs when an ideal solution is selected from the chosen issue among numerous elective solutions. A nature-inclined algorithm is a speculative technique to handle these enhancement issues. The proposed Feline storm algorithm develops by integrating the characteristic features of Feline and the mutual interaction between human beings as in [36]. The advantage of the hybrid model is that it avoids the intersection to the local optimum and provides minimum convergence time with low computational cost.

(i) Solution Encoding: The prime objective of the solution encoding process is to encode the entire data of the table into the isolated vector to enhance the privacy of the clinical data through anonymization. The semi-attribute age and zip code provide three different layers, where gender attribute consists of two stages. To implement the generalization, all the reports must be reposed in a single line. Information on patients' zip codes, gender, and age are illustrated in Figures 2 to 4. The level of the postal division is one; in figure 2, all the postal divisions of the structure 4453 are changed to 44*, and the story of the character's age is 2, thus figure 3, implying that the gender is male or female changes. In figure 4, the level of age is two like 20-30 to 30-40, while the zip

code and the gender. A similar degree of speculation is done for every one of the reports, and the protection of the reports depends on wellness. When each property is changed into the overall organization, wellness is determined in the reports. The arrangement encoding transforms the report into the summed-up report, and the security of this information is assessed and upgraded before distributing the information.

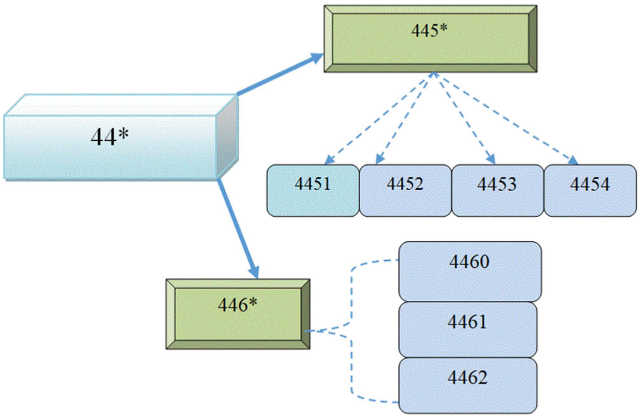


Figure 2. *Architecture of the patient's zip code.*

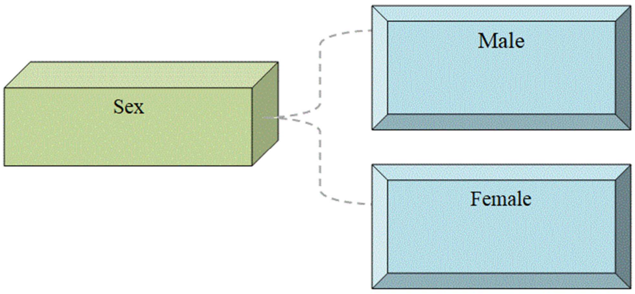


Figure 3. *Architecture of the patient's gender.*

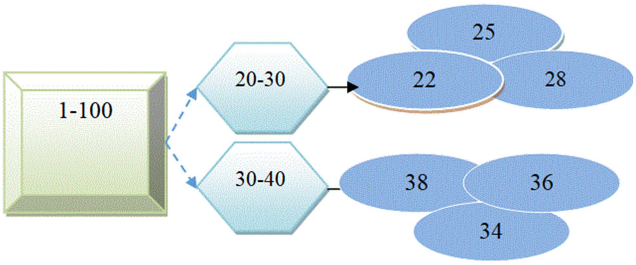


Figure 4. *Architecture of the Patients Age.*

(ii) Fitness Estimation: The fitness function is estimated to determine the best data, ensuring high privacy for the multi-institutional clinical data. The fitness function is determined by evaluating the generalized information loss $\mathrm{GIn}_{\mathrm{loss}}$ and the average equivalence class size $\mathrm{Class}_{\mathrm{avg}}$ the fitness values are estimated to escalate the privacy and reduce the utility once the privacy restriction is satisfied. The fitness attributes and the data utility are inversely proportional to each other. Hence, the loss of the function is maintained at a low level to obtain the maximum utility. The fitness restrictions are mathematically represented in the following expression.

$$f(P) = Co_1{}^{*}GIn_{loss}(P) + Co_2{}^{*}class_{avg}(P) \quad (4)$$

The privacy and the utility of data are supported through the minimum value of $Class_{avg}$alsoG ln_{loss} . The equation (5) is exposed to three conditions such as

$$\left.\begin{array}{l} k \geq r_d(D,a) \\ l \geq r_{def1}(D,a) \\ m \geq r_{def2}(D,a) \end{array}\right\} \quad (5)$$

The information is mathematically represented in the following equation:

$$GIn_{loss}(P) = \frac{1}{TR\times TQ} \times \sum_{\alpha=1}^{TQ} \sum_{\beta=1}^{TR} \frac{UB_{\alpha\beta}-LB_{\alpha\beta}}{UB_{\varepsilon}-LB_{\alpha}} \quad (6)$$

TR and TQ demonstrate the total number of records and quasi-identifier, respectively; UB represents the upper bound and LB denotes the lower bound.

$Class_{avg}$ is mathematically represented in the following equation,

$$Class_{avg} = \frac{TR}{|Iden_{sen}|^{*}k} \quad (7)$$

$Iden_{sen}$, demonstrates the sensitive identifiers. The function estimates the number of service providers. $r_{def1}(D,a), r_{def2}(D,a)$ determines the amount of the described sensitive attributes and $r_d(D,a)$ Shows the duplicate record.

3.2. Proposed Feline-Storm Algorithm

Felines appear to be sluggish and spend a more significant part of their time resting. Still, their awareness is exceptionally high during their rests, and they are conscious of what transpires around them. Thus, they continually notice the environmental factors brilliantly and intentionally; as soon as they see their objective, they move rapidly. On the other hand, humans can sort out the best solution for any problem through mutual interactions. Therefore, the proposed calculation is developed to integrate the two principal behaviors of felines and humans' mutual interaction. The proposed Feline-storm algorithm consists of three stages: idle phase, detecting phase, solution generating phase, and integrating step. A brief description of every phase is mentioned below.

a) Idle Phase: The idle phase imitates the idle characteristics of the Feline where four principal attributes play significant tasks: Exploring the memory pool (EMP), exploring the scope of the chosen dimension (ESD), estimating size to change (EDC), and self-position considering (SPC). Notably, SMP indicates the extent of exploring memory for felines. It characterizes several possible solutions from which the Feline will select the most optimal one. For example, if EMP was set to 5, then every single Feline generates five new solutions, one of which will be chosen to be the following situation of the Feline. The most effective method to randomize the new positions will rely upon the other two boundaries: EDC and ESD. EDC characterizes the number of measurements to alter within the range of$[0,1]$. As a case in point, if the search space has 5 measures and CDC is set to 0.2, then four irregular measurements out of the five should be altered at that point for each Feline, and the other one stays something very similar. ESD is the mutative proportion for the chosen measurements, i.e., it characterizes the measure of transformation and alterations for those measurements determined by EDC. SPC is a Boolean worth, which indicates whether the current situation will be elected as the probable solution. The steps involved in idle mode are.

a) Estimate the current position of the Feline.

b) Find out the new position of the Feline as shown in equation (8).

$$P_F^{t+1} = (1 + R^* ESD) * P_F^t \tag{8}$$

Where R represents the random number, P_F^{t+1} refers to a new position and P_F^t is the senior position of feline c) Estimate the fitness value for the possible solution as equation d) Based on the fitness value of the possible solution, select the best position.

b) Tracking Mode: This mode imitates the tracking characteristics of the Feline. For the first iterative process, the random velocity is assigned to all the dimensions of the Feline position. Yet, the velocity updates for the later steps. The steps involved in the tracking model are as;

a) Update velocities V_F for all dimensions as mentioned in equation (9).

$$V_F^{t+1} = V_F^t + R_1 C(P_F^{best} - P_F^t) \tag{9}$$

Where, V_F^t is the random velocity, R_1 denotes the random number within the range $[0,1]$ and C refers to the constant within the range$[0,1]$, andP_F^{best} is the best position of the Feline.

b) If the velocity exceeds the extreme value, the velocity value is considered maximum velocity.

c) Update the position of the Feline as mentioned in equation (9).

$$P_F^{t+1} = P_F^t + V_F^{t+1} \tag{10}$$

Substitute equation (9) in (10)

$$P_F^{t+1} = P_F^t + V_F^t + R_1 C(P_F^{best} - P_F^t) \tag{11}$$

Hence, it can be inferred that equation (11) provides the best solution for the Feline. This solution is integrated with the solution provided by the humans during the mutual interaction process.

c) Solution Generating Phase: As every individual may have encountered when they deal with a troublesome issue, a gathering of people, particularly with various backgrounds, get together to conceptualize the problem and can generally address the solution with high probability. Thus, in mutual interaction, people from different backgrounds will generate other ideas, and the idea that possesses the best solution is selected.

d) Population Generation Phase. In the population generation phase, the X number of solutions are randomly generated. a) Congregate the X solutions into Y clusters. b) Estimate the X solutions. c) Rank each solution in the cluster and note down the best solution. d) Randomly create a value within the range$[0,1]$, e) Estimate the new solution. The newly estimated solution compares with the old solution; if the old solution is better, the new solution is replaced by the best solution. If the new solution is better, it remains the updated solution. There are four directives involved in the mutual interaction process. The coordinator is not involved in creating the new ideas but initiating the interaction process. The only requirement of selecting the coordinator is the facilitation experience but less expertise on the background knowledge related to the problem to be solved as possible.

The four directives are 1) Intermitted Judgment, 2) There Are No Restrictions, 3) Cross-Fertilizer, and 4) Go for Quantity. Directive 1 reveals that all the ideas attained by the mutual interaction process are considered good; no statements should be treated as wrong. It is unwise to judge whether a proposed idea is a good or bad idea. Any judgment or criticism must be held back until at least the end of the brainstorming process. The second directive demonstrates that anything that strikes the mind during the mutual interaction is valuable to be distributed and recorded. Directive 3 indicates that numerous ideas should be generated concerning the already developed ideas. The already generated ideas serve as hints to create new ideas. Directive 4 demonstrates that various ideas should be generated. It is said that the increase in quantity will increase the quality of the pictures. Developing a vast quantity of views will improve the quality of the generated ideas.

Steps Involved in Mutual Interaction Process: a) Gather all the individuals with diverse backgrounds for the mutual interaction process. b) Create a possible solution. c) Consider 3 to 5 individuals as the observer to select the best ideas to solve the issues. d) While utilizing the selected ideas as the primary solution and others as clues, generate other ideas. e) Repeat the same procedures until it attains maximum iteration. The best solution thus generated by the mutual interaction process is formulated in equation (12).

$$P_{mi}^{t+1} = P_{mi}^{sel} + G^* X(\mu, \sigma) \tag{12}$$

Where, P_{mi}^{sel} Refers to the solution selected in the mutual interaction to generate new ideas or solutions.Grepresents the Gaussian random variable, $X(\mu,\sigma)$ Denotes the Gaussian random function with variance $\sigma$and meanμ.

f) Integrating phase: The best solution is obtained by integrating the best solution obtained from the Feline and mutual interaction process to get the optimal solution.

$$P_{fs}^{best} = 0.5[P_F^{t+1}] + 0.5[P_{mi}^{t+1}] \tag{13}$$

Substituting (11) and (12) in (13) we get

$$P_{fs}^{best} = 0.5[P_F^t + V_F^t + R_1 C(P_F^{best} - P_F^t)] + 0.5\left[P_{mi}^{sel} + G * X(\mu, \sigma)\right] \tag{14}$$

The equation (14) is the best solution obtained by the feline-storm optimization algorithm, which is generated by integrating the Feline's characteristics and the mutual interaction of human beings. It provides the best solution as it combines the advantages of Feline and Human beings. The privacy preservation technique utilizes the hybrid algorithm to emphasize exploitation or the exploration phase at any instance, avoiding the local optima and pre-mature convergence.

4. Result and Discussion

The Feline-Storm algorithm is used to enhance the Privacy Preservation Technique for the multi-institutional clinical data, and the model's effectiveness is assessed concerning other approaches.

4.1. Experimental Setup

To assess the model's performance and advancement, it is implemented in MATLAB tool 2020 on Windows 10 with 8GB RAM.

4.2. Dataset Description

The standard Healthcare datasets, Cleveland, Hungary, Switzerland, and VA database considered from the UCI machine learning repository will be utilized for the research. This dataset contains four databases regarding cardiac disease diagnosis. The data are collected from the four locations: Cleveland Clinical Foundation, Hungarian Institute of Cardiology in Budapest, Veterans Administration - Long Beach Medical Center in California, University Hospital, and Zurich, Switzerland. It comprised a total of 76 numeric-valued attributes, of which 14 attributes such as age, sex, cp (chest pain type, which 1 denotes typical angina, 2 denotes atypical angina, 3 denotes non-anginal pain and 4 asymptotic), trestbps (resting blood pressure in mmHg), chol (serum cholesterol), FBS (fasting blood sugar), restecg (resting ECG), thalach (maximum heart rate achieved), exang (exercise induced angina), old peak (ST depression caused by exercise relative to rest), slope (the slope of the peak exercise ST segment), ca (number of significant vessels colored by flourosopy), thal (thalassemia), and num (diagnosis of heart disease) are used for the evaluation process. Generally, the Cleveland database is widely utilized by ML researchers as it provides high accuracy with previous experiments. The primary purpose of creating this database is to determine the cardiac disease in the patients. The sensitive information, such as social security numbers and names, is replaced with dummy values. Table 1 denotes the class distribution of each dataset.

4.3. Performance Metrics

The key parametrics, such as generalized information loss, Average equivalence class size, and fitness parameters, are utilized for the performance evaluation and the comparative analysis of the proposed methodology. The brief description is described in the following section.

a) Generalized Information Loss(GIn_{loss})*:* The GIn_{loss}Is the prime attribute utilized to determine the fitness function of the system. An effective privacy-preserving technique should provide the minimum value of information loss. The $GIn_{loss}$is mathematically expressed in the following equation

$$GIn_{loss} = \frac{1}{TR \times TQ} \times \sum_{\alpha=1}^{TQ} \sum_{\beta=1}^{TR} \frac{UB_{\alpha\beta} - LB_{\alpha\beta}}{UB_{\alpha} - LB_{\alpha}} \tag{15}$$

The $TRandTQ$ demonstrates the total number of records and quasi-identifiers respectively; theUB represents the upper and LB represents the lower bound.

b) Average Equivalence Class Size Metric$\left(class_{avg}\right)$*:* An effective privacy preservation technique maintains the minimum value of $class_{avg}$, and it is mathematically represented in the following equation

$$Class_{avg} = \frac{TR}{|Iden_{sen}|^{*}k} \tag{16}$$

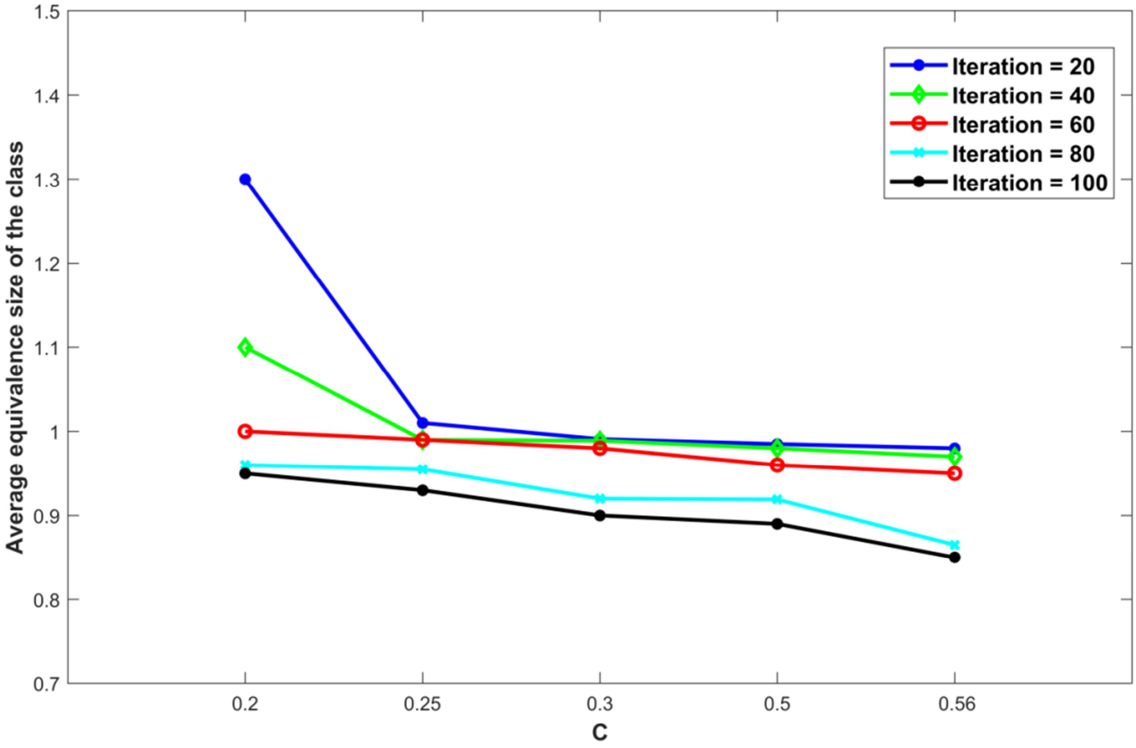
The $iden_{sen}$ demonstrates the sensitive identifiers.

c) Fitness function: The fitness score acquired from the record will be maintained at a minimum to determine the privacy of the record

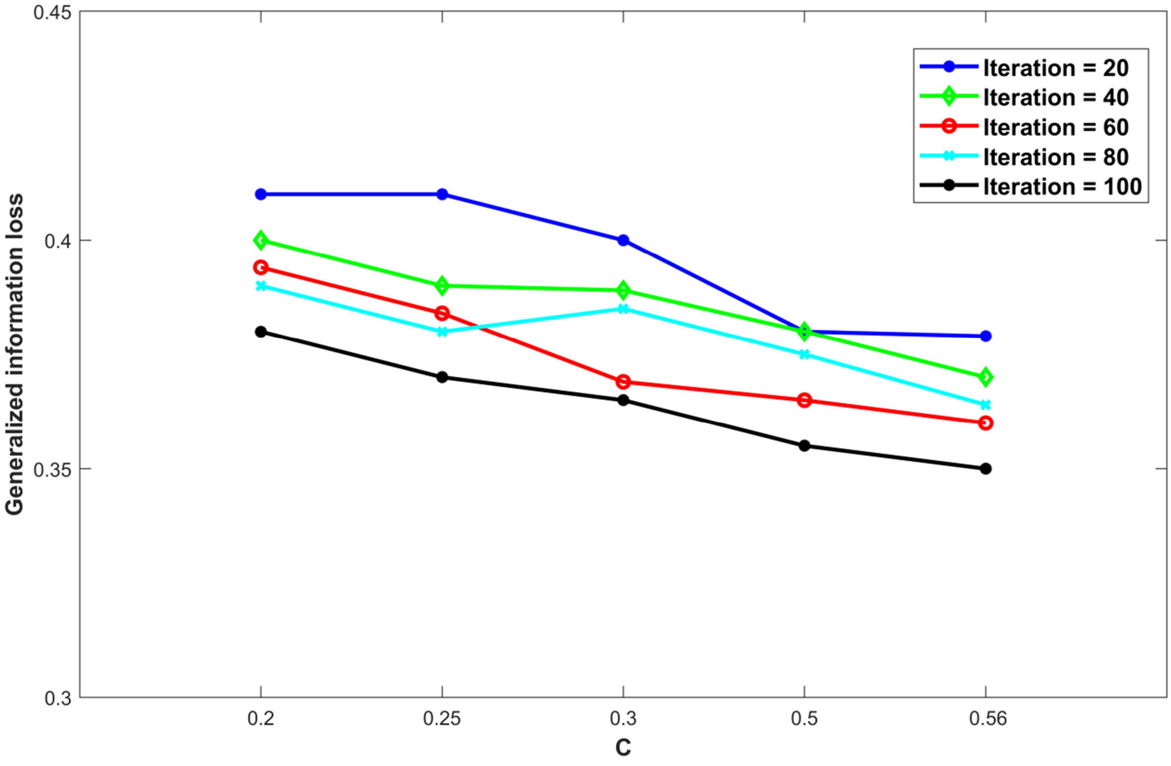
$$Fit = Con_1 * GIn_{loss} + Con_2 * class_{avg} \tag{17}$$

4.4. Performance Analysis

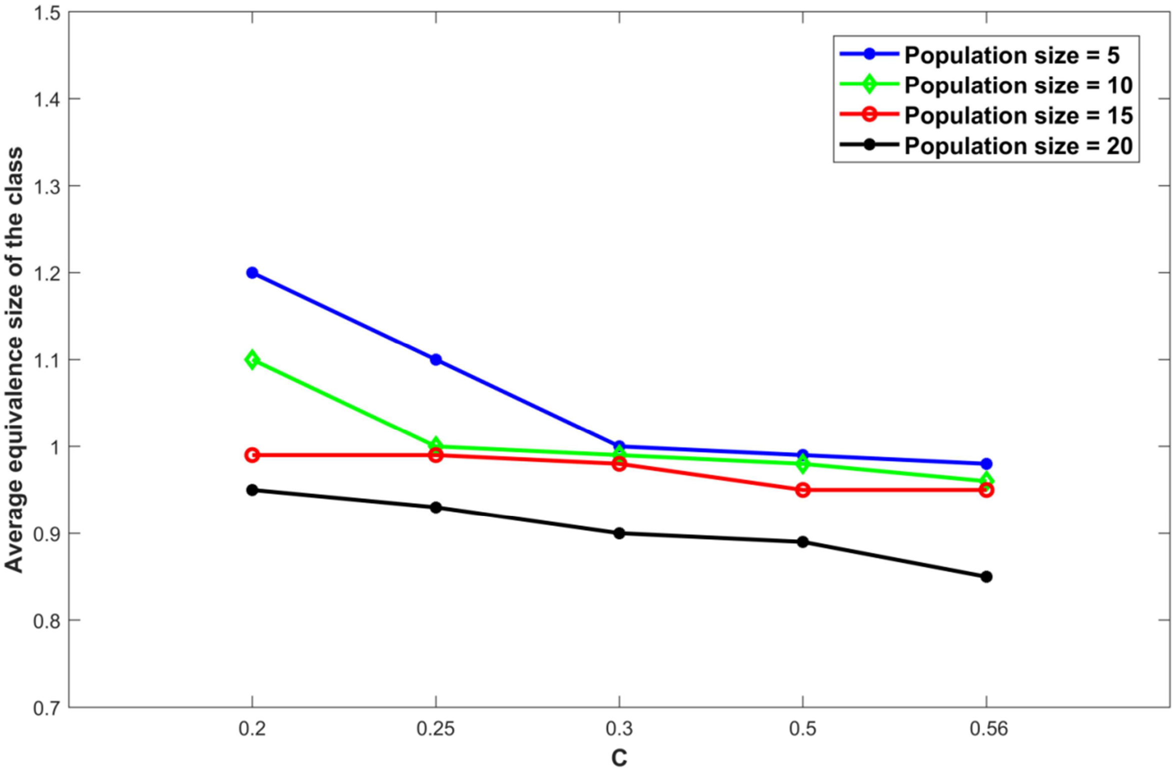
The performance obtained by the proposed Feline-Storm optimization regarding the class average size concerning the C-mixture value is depicted in Figure 5a). At the 100th iteration the minimum $Class_{avg}$is achieved when C-mixture= 0.56. This demonstrates that the Feline-Storm based privacy preservation technique provides better performance in terms of $Class_{avg}$at 100th iteration and at the C-mixture of 0.56. The performance analysis in terms of $GIn_{loss}$is illustrated in Figure 5b). At 100th iteration the GIn_{loss} attained by the proposed Feline-Storm Based Privacy Preservation Technique are 0.38, 0.37, 0.365, 0.355 and 0.35 for the c-mixture value of 0.2, 0.25, 0.3, 0.5 and 0.56 respectively. It is demonstrated that the minimum information loss of 0.35 is obtained at C-mixture =0.56 and at the 100th iteration. The minimum information loss is achieved due to the optimal selection of the C-mixture value using the proposed Feline-Storm-based privacy preservation model. Figure 5c) shows the performance evaluation of the proposed Feline-Storm Based Privacy Preservation Technique in terms of $Class_{avg}$concerning the C-mixture is demonstrated in the figure below. When the population size is 15, the $Class_{avg}$ attained by the proposed privacy preservation method is found to be 0.99, 0.99, 0.98, 0.95 and 0.95 respectively. The $Class_{avg}$ attained by the proposed Feline-Storm Based Privacy Preservation Technique are found to be 0.95, 0.93, 0.9, 0.89 and 0.85 for the C-mixture value of 0.2, 0.25, 0.3, 0.5 and 0.56, respectively. It is illustrated that the $Class_{avg}$reduces with an increase in the C-mixture value and the population size. The minimum class average equivalence size is obtained by determining the optimal C-mixture value by the proposed Feline-Storm optimization algorithm. Figure 5d) shows the performance analysis of the proposed Feline-Storm Based Privacy Preservation Technique in terms of GIn_{loss}concerning the C-mixture value based on the number of populations. It states that the proposed privacy preservation method attains the minimum loss when the C-mixture value is 0.56, and the population size is 20. The minimum generalized information loss is obtained as the best-fitted C-mixture value is optimally selected by proposed Feline-storm optimization algorithm.



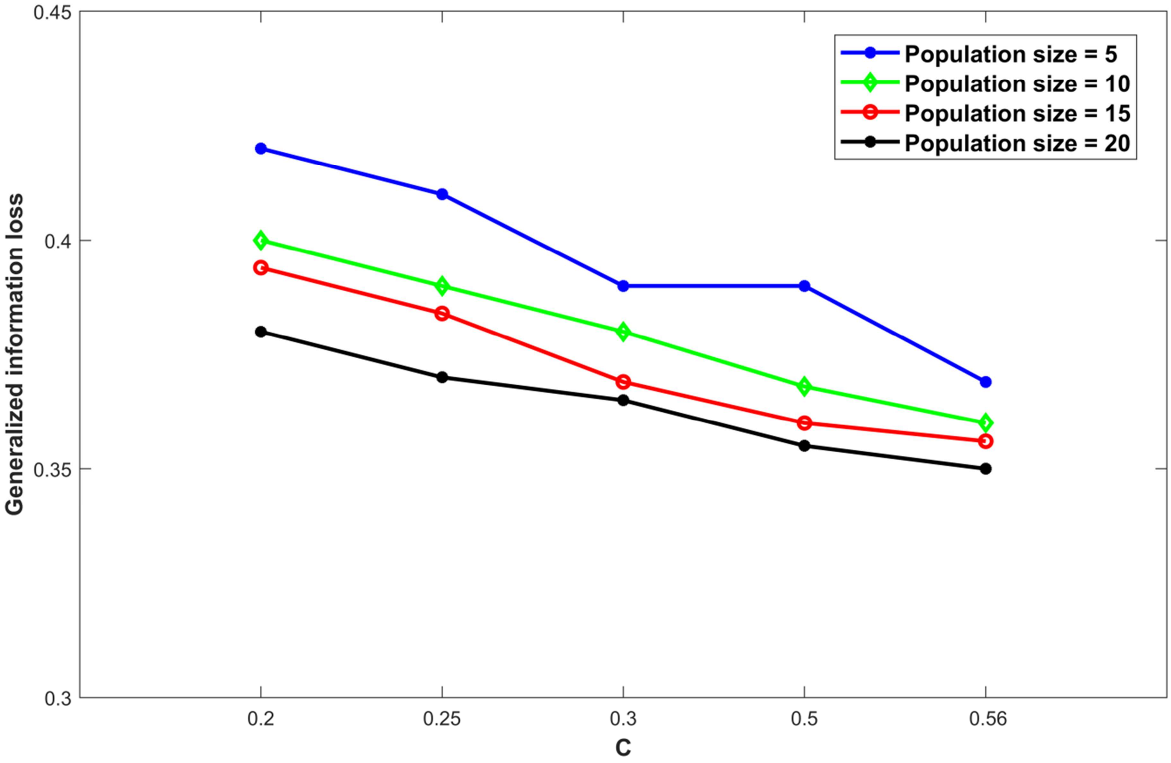
a) Class$_{avg}$



b) Gn$_{loss}$



c) $\text{Class}_{\text{avg}}$

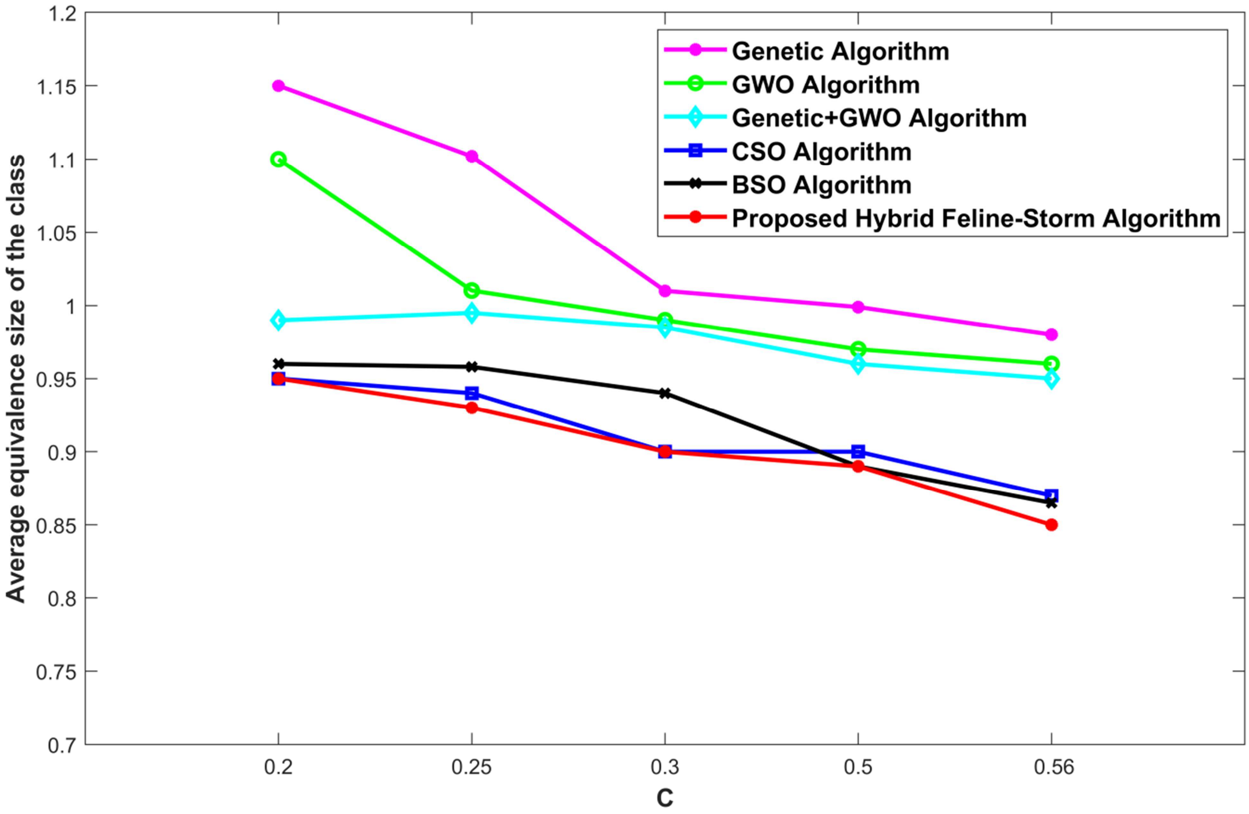


d) Gn_{loss}

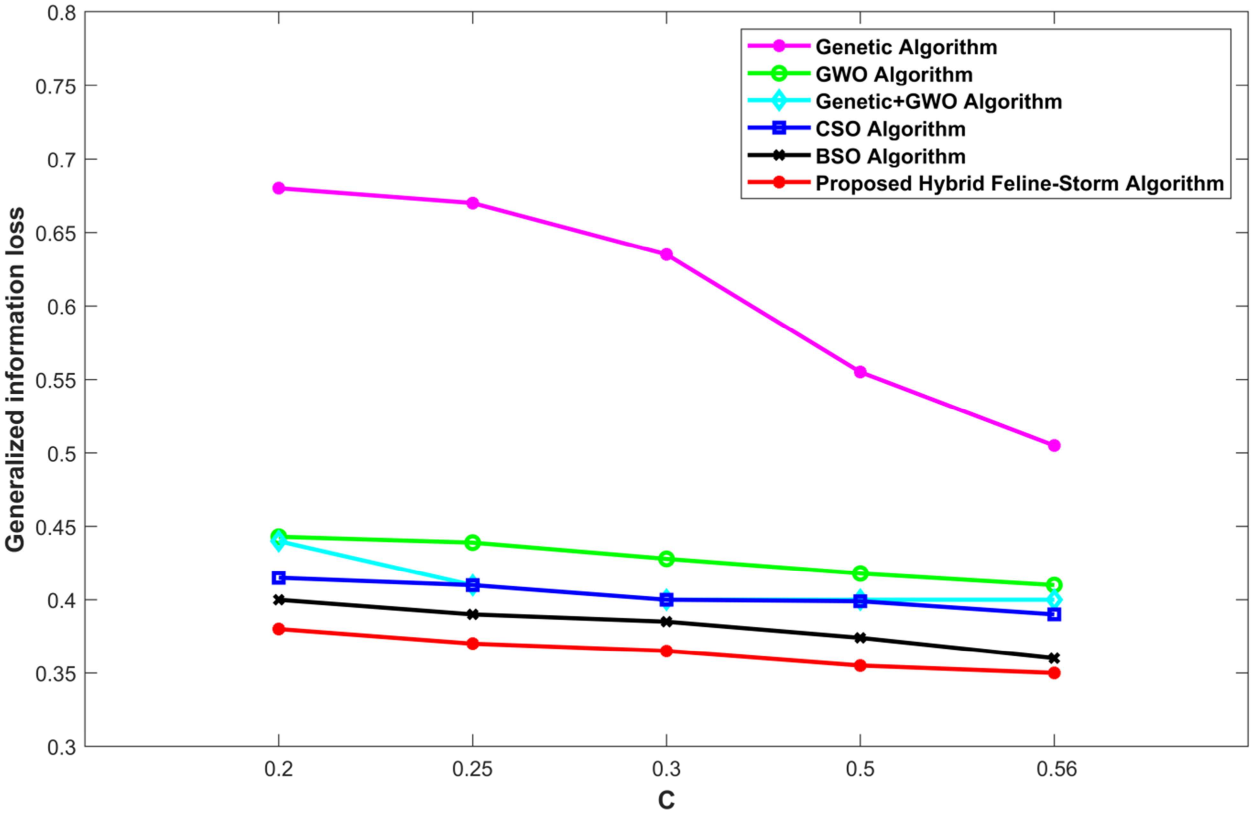
Figure 5. *Performance analysis based on the iteration and population size a)* $class_{avg}$ *b)* GIn_{loss} *c)* $class_{avg}$ *d)* GIn_{loss}.

4.5. Comparative Methods

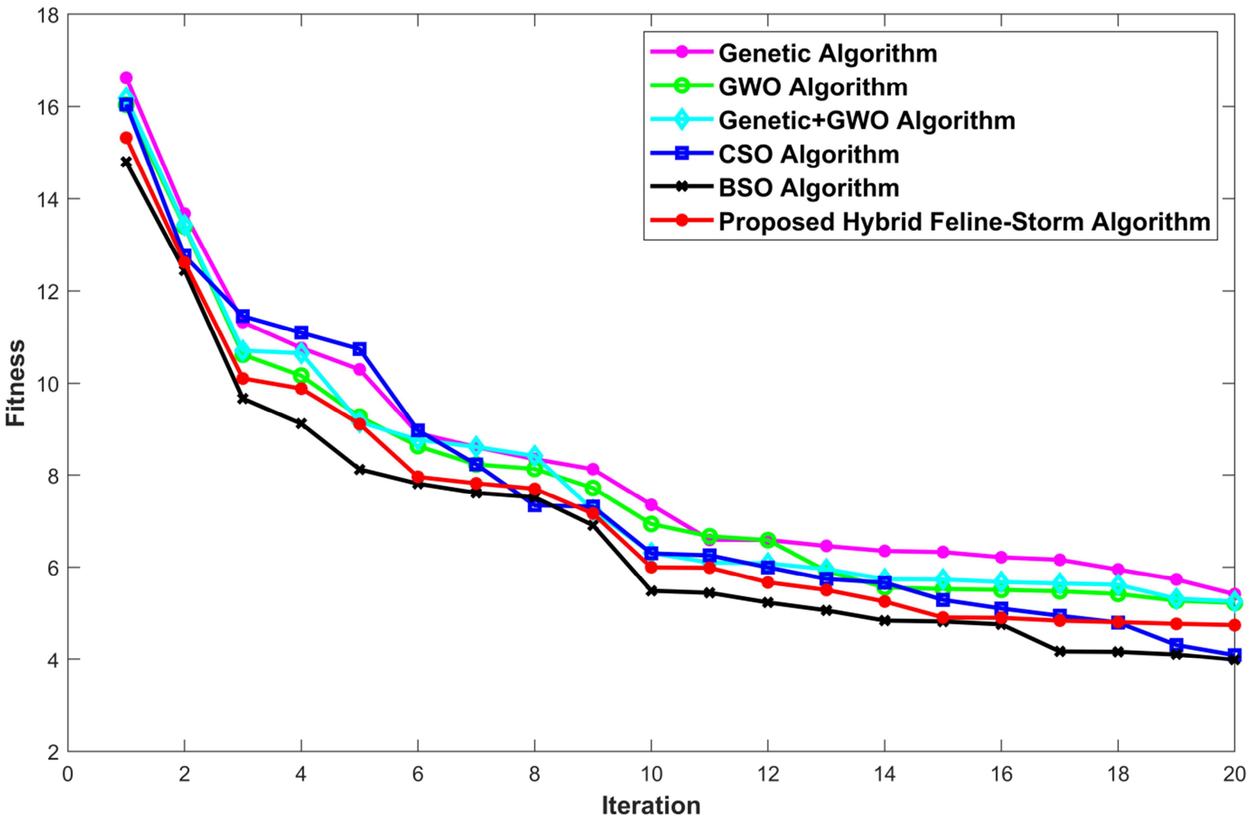
Some of the conventional methods, such as a) Genetic algorithm [36], b) GWO algorithm [37], c) Genetic +Grey Wolf Optimization (GWO) algorithm [38], Cat Swarm Optimization algorithm (CSO) [39], and d) Brainstorm Optimization BSO [40] algorithm is utilized for the comparative evaluation.



a) Class_{avg}



b) Gn_{loss}



a) Fitness

Figure 6. *Comparative analysis based on a) $class_{avg}$ b) GIn_{loss} c) Fitness.*

(i) Comparative evaluation in terms of $Class_{avg}$: Figure 6a) shows the comparative analysis of the proposed Feline-Storm Based Privacy Preservation Technique in terms of $Class_{avg}$. For the C-mixture value of 0.56, the $Class_{avg}$ attained by the comparative methods such as BSO algorithm are found to be 0.89. In contrast, the proposed Hybrid Feline-Storm algorithm is found to be 0.86. This demonstrates that the proposed Feline-Storm-based privacy preservation techniques attain the lowest $Class_{avg}$ compared to all the other competent methods.

(ii) Comparative evaluation in terms of GIn_{loss} : Figure 6b) shows the comparative analysis of the proposed Feline-Storm Based Privacy Preservation Technique with conventional methods in terms of GIn_{loss} . The GIn_{loss} of the proposed Feline-Storm algorithm at a C-mixture value of 0.2 is found to be 0.38. The GIn_{loss} attained by the conventional Genetic algorithm, GWO algorithm, Genetic + GWO algorithm, CSO algorithm BSO algorithm, and proposed Feline-Storm algorithm are found to be 0.56, 0.535, 0.45, 0.43, 0.4, and 0.38, respectively.

(iii) Comparative evaluation of Fitness Fit: The Fitness score obtained by the conventional methods, such as BSO algorithm at 2nd iteration, is found to be 14.8008. The proposed Feline-Storm Based Privacy Preservation model attains a fitness score of 15.3248. At 10th iteration, the fitness score achieved by the BSO algorithm is 5.4856. Figure 6c) shows the comparative analysis of the proposed Feline-Storm Based Privacy Preservation in terms of fitness value, showing a fitness score of 5.9950, which is comparatively higher than the conventional model. The BSO algorithm attains the fitness score of 3.9918 at the 2020th iteration. The proposed Feline-Storm Based Privacy Preservation model reaches the fitness score of 20th iteration of about 4.7457.

4.6. Comparative Discussion

This section discusses the evaluation of the Feline-Storm Based Privacy Preservation model in terms of GIn_{loss} , $Class_{avg}$ and Fit . The minimum of GIn_{loss} , $Class_{avg}$ and Fit attained by various conventional methods along with the proposed Feline-Storm algorithm, is illustrated in the table. Thus, the proposed Hybrid Feline-Storm algorithm exceeds all the conventional methods in terms of GIn_{loss} , $Class_{avg}$ and Fit .

Table 1. *Comparison Result.*

Methods	$class_{avg}$	GIn_{loss}	**Fitness score**
Genetic Algorithm	1.03	0.56	5.4118
GWO algorithm	1	0.535	5.2233
Genetic +GWO algorithm	0.95	0.45	5.2564
CSO algorithm	0.91	0.43	4.0936
BSO algorithm	0.9	0.4	3.9918
Proposed Hybrid Feline-Storm algorithm	0.85	0.38	4.7457

5. Conclusion

In this research, the effectiveness of the proposed Hybrid Feline-Storm algorithm-based privacy preservation techniques compared to the competent methods such as the genetic algorithm, GWO algorithm, Genetic + GWO algorithm, CSO algorithm, and BSO algorithm. The implementation is carried out in MATLAB. The UCI machine learning repository heart disease dataset is used for the proposed methodology's performance and comparative analysis. The experimental outcome demonstrates that the proposed Hybrid Feline-Storm algorithm-based privacy preservation techniques attain the lowest information loss, class average size and fitness measure compared to all the competent methods. The information loss, class average size, and fitness measure attained by the proposed methodology are found to be 0.85, 0.38, and 4.7457, respectively.

Ethics Statements

The relevant informed consent was obtained from those subjects from open sources.

Credit Author Statement

Sagarkumar Patel conceived the presented idea and designed the analysis. Also, he carried out the experiment and wrote the manuscript with support from Rachna Patel. All authors discussed the results and contributed to the final manuscript. All authors read and approved the final manuscript.

Declaration of Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

I want to express my very great appreciation to the co-authors of this manuscript for their valuable and constructive suggestions during the planning and development of this research work.

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Liu, Yi, J. Q. James, Jiawen Kang, Dusit Niyato, and Shuyu Zhang. "Privacy-preserving traffic flow prediction: A federated learning approach." IEEE Internet of Things Journal 7, no. 8 (2020): 7751-7763.

[2] Zhao, Lingchen, Qian Wang, Qin Zou, Yan Zhang, and Yanjiao Chen. "Privacy-preserving collaborative deep learning with unreliable participants." IEEE Transactions on Information Forensics and Security 15 (2019): 1486-1500.

[3] Lu, Yunlong, Xiaohong Huang, Yueyue Dai, Sabita Maharjan, and Yan Zhang. "Blockchain and federated learning for privacy-preserved data sharing in industrial IoT." IEEE Transactions on Industrial Informatics 16, no. 6 (2019): 4177-4186.

[4] Wang, X., Zhang, A., Xie, X., & Ye, X. (2019). Secure‐aware and privacy‐preserving electronic health record searching in cloud environment. International Journal of Communication Systems, 32(8), e3925.

[5] Wang, Y., Zhang, A., Zhang, P., & Wang, H. (2019). Cloud-assisted EHR sharing with security and privacy preservation via consortium blockchain. IEEE Access, 7, 136704-136719.

[6] Wilkowska, W., & Ziefle, M. (2012). Privacy and data security in E-health: Requirements from the user's perspective. Health Informatics Journal, 18(3), 191-201. doi: 10.1177/1460458212442933.

[7] Makhdoom, Imran, Ian Zhou, Mehran Abolhasan, Justin Lipman, and Wei Ni. "PrivySharing: A blockchain-based framework for privacy-preserving and secure data sharing in smart cities." Computers & Security 88 (2020): 101653.

[8] Kaissis, Georgios, Alexander Ziller, Jonathan Passerat-Palmbach, ThéoRyffel, DmitriiUsynin, Andrew Trask, Ionésio Lima Jr et al. "End-to-end privacy preserving deep learning on multi-institutional medical imaging." Nature Machine Intelligence 3, no. 6 (2021): 473-484.

[9] Sheller, M. J., Reina, G. A., Edwards, B., Martin, J., & Bakas, S. (2019). Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation. In Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4 (pp. 92-104). Springer International Publishing.

[10] Sarrab, M., & Alshohoumi, F. (2020). Privacy Concerns in IoT a Deeper Insight into Privacy Concerns in IoT Based Healthcare. International Journal of Computing and Digital Systems, 9 (03). http://dx.doi.org/10.12785/ijcds/090306.

[11] Xi, W., & Ling, L. (2016, December). Research on IoT privacy security risks. 2016 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), pp. 259-262, IEEE. doi: 10.1109/iciicii.2016.0069.

[12] Xu, H., Russell, T., Coposky, J., Rajasekar, A., Moore, R., de Torcy, A., Wan, M., Shroeder, W and Chen, S. Y. (2017). iRODS primer 2: integrated rule-oriented data system. Synthesis Lectures on Information Concepts, Retrieval, and Services, 9 (3), 1-131.

[13] Seliem, M., Elgazzar, K., & Khalil, K. (2018). Towards privacy preserving IoT environments: a survey. Wireless Communications and Mobile Computing. doi: 10.1155/2018/1032761.

[14] Shahzad, A., Lee, Y. S., Lee, M., Kim, Y. G., & Xiong, N. (2018). Real-time cloud-based health tracking and monitoring system in designed boundary for cardiology patients. Journal of Sensors, 320278, 15.

[15] Sharma, S., Chen, K., & Sheth, A. (2018). Toward practical privacy-preserving analytics for IoT and cloud-based healthcare systems. IEEE Internet Computing, 22 (2), 42-51.

[16] Shi, Y. (2011). Brainstorm Optimization Algorithm. International Conference in Swarm Intelligence, 303-309. doi: 10.1007/978-3-642-21515-5_36.

[17] Siddiqa, A., Hashem, I. A. T., Yaqoob, I., Marjani, M., Shamshirband, S., Gani, A., & Nasaruddin, F. (2016). A survey of big data management: Taxonomy and state-of-the-art. Journal of Network and Computer Applications, 71, 151-166. doi: 10.1016/j.jnca.2016.04.008.

[18] Lexchin, Joel. "Those who have the gold make the evidence: how the pharmaceutical industry biases the outcomes of clinical trials of medications." Science and engineering ethics 18 (2012): 247-261.

[19] Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. Journal of Business Research, 70, 263-286. doi: 10.1016/j.jbusres.2016.08.001.

[20] Solangi, Z. A., Solangi, Y. A., Chandio, S., bin Hamzah, M. S., & Shah, A. (2018, May). The future of data privacy and security concerns in Internet of Things. 2018 IEEE International Conference on Innovative Research and Development (ICIRD), 1-4, IEEE. doi: 10.1109/ICIRD.2018.8376320.

[21] Sonune, S., Kalbande, D., Yeole, A., & Oak, S. (2017, June). Issues in IoT healthcare platforms: A critical study and review. 2017 International Conference on Intelligent Computing and Control (I2C2), 1-5, IEEE.

[22] Stergiou, C., Psannis, K. E., Gupta, B. B., & Ishibashi, Y. (2018). Security, privacy & efficiency of sustainable cloud computing for big data & IoT. Sustainable Computing: Informatics and Systems, 19, 174-184. https://doi.org/10.1016/j.suscom.2018.06.003.

[23] Stojkov, M., Sladić, G., Milosavljević, B., Zarić, M., & Simić, M. (2019). Privacy concerns in IoT smart healthcare system. Conference: 8th International Conference on Information Society and Technology (ICIST), 1.

[24] Sudhakar, R. V., & Rao, T. C. M. (2020). Security aware index based quasi–identifier approach for privacy preservation of data sets for cloud applications. Cluster Computing, 1-11.

[25] Suneetha, V., Suresh, S., & Jhananie, V. (2020). A novel framework using Apache spark for privacy preservation of healthcare big data. Proceedings of 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 743-749, IEEE.

[26] Sweeney, L. (2002). Achieving k-anonymity privacy protection using generalization and suppression. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10 (05), 571-588.

[27] Tabassum, K., Ibrahim, A., & El Rahman, S. A. (2019, April). Security issues and challenges in IoT. 2019 International Conference on Computer and Information Sciences (ICCIS), 1-5. doi: 10.1109/ICCISci.2019.8716460.

[28] Tariq, N., Asim, M., Al-Obeidat, F., Zubair Farooqi, M., Baker, T., Hammoudeh, M., & Ghafir, I. (2019). The security of big data in fog enabled IoT applications including blockchain: A survey. Sensors, 19 (8), 1788.

[29] Tawalbeh, L. A., Muheidat, F., Tawalbeh, M., & Quwaider, M. (2020). IoT Privacy and security: Challenges and solutions. Applied Sciences, 10 (12), 4102.doi: 10.3390/app10124102.

[30] Terzi, D. S., Terzi, R., & Sagiroglu, S. (2015). A survey on security and privacy issues in big data. Proceedings of 10th International Conference for Internet Technology and Secured Transactions (ICITST), 202-207. doi: 10.1109/ICITST.2015.7412089.

[31] Truta, T. M., & Vinay, B. (2006). Privacy protection: p-sensitive k-anonymity property. Proceedings of IEEE International Conference on Data Engineering Workshops (ICDEW'06), 94-94. doi: 10.1109/ICDEW.2006.116.

[32] Tucker, K., Branson, J., Dilleen, M., Hollis, S., Loughlin, P., Nixon, M. J., & Williams, Z. (2016). Protecting patient privacy when sharing patient-level data from clinical trials. BMC Medical Research Methodology, 16(1), 5-14. https://doi.org/10.1186/s12874-016-0169-4.

[33] Uddin, M. A., Stranieri, A., Gondal, I., & Balasubramanian, V. (2018). Continuous patient monitoring with a patient centric agent: A block architecture. IEEE Access, 6, 32700-32726. doi: 10.1109/ACCESS.2018.2846779.

[34] Wachter, S. (2018). Normative challenges of identification on the Internet of Things: Privacy, profiling, discrimination, and the GDPR. Computer Law & Security Review, 34 (3), 436-449. https://doi.org/10.1016/j.clsr.2018.02.002.

[35] Li, Xiaoxiao, Yufeng Gu, NichaDvornek, Lawrence H. Staib, Pamela Ventola, and James S. Duncan. "Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: ABIDE results." Medical Image Analysis 65 (2020): 101765.

[36] Mirjalili, Seyedali, and SeyedaliMirjalili. "Genetic algorithm." Evolutionary Algorithms and Neural Networks: Theory and Applications (2019): 43-55.

[37] Rezaei, Hossein, Omid Bozorg-Haddad, and Xuefeng Chu. "Grey wolf optimization (GWO) algorithm." Advanced optimization by nature-inspired algorithms (2018): 81-91.

[38] Zhou, Jian, Shuai Huang, Tao Zhou, Danial JahedArmaghani, and Yingui Qiu. "Employing a genetic algorithm and grey wolf optimizer for optimizing RF models to evaluate soil liquefaction potential." Artificial Intelligence Review 55, no. 7 (2022): 5673-5705.

[39] Chu, Shu-Chuan, Pei-Wei Tsai, and Jeng-Shyang Pan. "Cat swarm optimization." In PRICAI 2006: Trends in Artificial Intelligence: 9th Pacific Rim International Conference on Artificial Intelligence Guilin, China, August 7-11, 2006 Proceedings 9, pp. 854-858. Springer Berlin Heidelberg, 2006.

[40] Shi, Yuhui. "An optimization algorithm based on brainstorming process." In Emerging Research on Swarm Intelligence and Algorithm Optimization, pp. 1-35. IGI Global, 2015.