[1] D Kavitha
[2] Dr. T. Adilakshmi
[3] Dr. M. Chandra Mohan

Preserving Privacy of IoT Healthcare Data using Differential Privacy and LSTM

Abstract: - The Internet of Things (IoT) is a powerful technology creating revolutions in multiple industries for ex: Traffic and Healthcare domains. The patient data collected by continuous monitoring using IoT will support in treating the patients and make a positive impact on patients' well-being and increase the efficiency of healthcare workers. It is crucial to be aware of certain drawbacks and risks associated with protecting the privacy of the patient data which is one of the major problems being faced in the healthcare domain. Harmful individuals/Agencies will use IoT devices to obtain private data of patients. It's of prime importance to protect privacy in healthcare. To improve the privacy of IoT Healthcare data, Geometric data perturbation along with Noise addition is introduced in this study utilizing Laplace Noise which comes under the framework of Differential Privacy. To increase accuracy, a deep learning technique Long Short-Term Memory (LSTM) is applied in this paper. LSTM has proven to be a superior model in accuracy when compared with other models like Decision Tree and Naive Bayes.

Keywords: Internet of Things (IoT); Long Short-Term Memory (LSTM); Laplace noise; IoT Healthcare Data; Differential Privacy.

1. Introduction

Data mining involves gathering vast quantities of accurate personal information, which is subsequently analysed. Among the many types of information that fall under this category are shopping patterns, medical histories, credit records and many more.

Having access to such information is a valuable resource for businesses and governments because it aids in decision-making and offers social advantages like better health care, traffic prediction, crime prevention, and national security. On the other hand, there is a huge risk and challenge that governments and organizations are facing in ensuring the privacy of the data is secured, and not falling into the wrong hands and misuse of the data. As data mining algorithms can extract sensitive information from unclassified data, there is a high possibility that individuals' privacy could be compromised [2]. Since individuals are ignorant of the utilization of data mining "behind the scenes," the breach of their privacy that results from secondary uses of data is even more serious. [3]

The challenging problem is: How can we satisfy the requirements of government bodies/institutions to frame guidelines and advance social/organizational/institutional objectives while guarding against misuse of the information received?

With the proliferation of internet-connected devices, social media platforms, IoT devices, and mobile applications a huge amount of data is being generated than ever before and being transferred across multiple networks consisting of sensitive information.

With so much data available if integrated well, with the right technologies like Data analytics and AI, there is a great opportunity for analysis of the data and use it for the betterment of larger society.

However, one of the common challenges faced in the process of data preservation is maintaining a fine balance between Privacy and Data Utility, as there is a challenge in the efficient retrieval of data and accuracy, in the process of improving privacy.

There have been numerous methods proposed earlier in data privacy preserving which have greatly advanced the field of data privacy-preserving and there is a growing need for finding efficient ways in data preserving that are suitable to the current digital world where security is not at the last mile but at every stage of the data pipeline, compliant with privacy regulations and standards, fast, less resource intensive, scalable and adaptive to different kinds of data.

Techniques that have been identified in the past and proven effective in real-world applications and integrate well with groundbreaking latest technologies like machine learning and AI.

The current need of the hour is to identify a solution that will provide a privacy guarantee to the sensitive data generated across multiple sources and also leverage current groundbreaking technologies like Machine Learning

[1] Assistant Professor, Dept of CSIT, IARE, JNTUH, Hyderabad, India. Email: kavithadasari.it2005@gmail.com
[2] Prof. & Head, Dept of CSE, Vasavi College of Engineering, Hyderabad, India. Email: hodcse@staff.vce.ac.in
[3] Professor of CSE & Principal of UCERS, Hyderabad, India. Email: c_miryala@jntuh.ac.in

and Artificial Intelligence.

Considering the above factors, extensive exploration was done on multiple privacy-preserving methods Et al. [58,59,60,61,62,63,64 & 65], of which Geometric data transformations (GDTM) and Differential Privacy (DP) methods have been identified, for below mentioned reasons.

GDTM methods can be applied to data of different natures and scales, efficient in providing data privacy at edge level in the cloud, but they do have their challenges like loss of information if not done right.

Differential Privacy provides a finer level of control to the data by usage of techniques like noise magnitude and has been extensively used in real-world applications.

We have amalgamated the above two methodologies into one, to enhance data privacy. As cited earlier, in the process of enhancing data privacy there is a risk of a decrease in data accuracy which calls for identifying methodologies to increase the accuracy of data retrieval, flexible, versatile, effective in handling noisy data, and delivering state-of-the-art performance.

After extensive analysis by Et al. [54,55,56,57], LSTM (Long short-term Memory) a deep learning technique has been identified for improving data accuracy, which not only meets the criteria mentioned but also addresses the vanishing gradient problem faced from traditional traditional RNNs and helps in efficient data accuracy.

In this research, a multidimensional geometric data perturbation method along with an optimized noise addition is proposed to increase privacy and a deep learning technique for improving privacy. The proposed technique can be employed for numerous types of popular data mining methods.

The block diagram of the processes involved in the proposed research is shown in Figure 1.

The current paper has been organized into the following stages:

- Stage 2 focuses on the works based on privacy protection from the previous research,
- Stage 3 proposes the methodology,
- Stage 4 summarizes the results obtained, and
- Stage 5 conclusion of the current paper research and its findings.

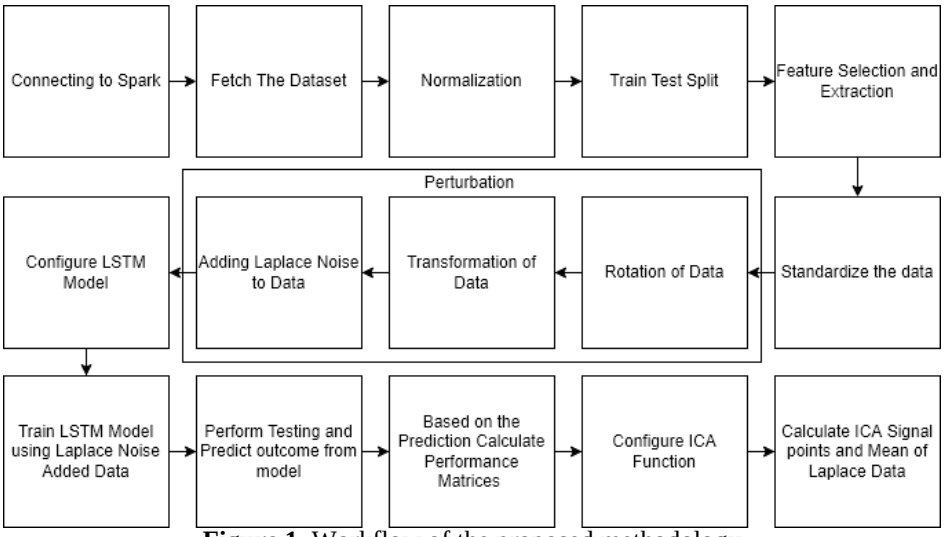

Figure 1. Workflow of the proposed methodology

2. Literature Review

When it comes to protecting user privacy, X. Zhang et al. [7] developed an anonymization method that is multidimensional and scalable based on MapReduce. In this instance, a scalable median-finding approach to incorporate the histogram technique for tuning the recursion's fineness was used. Using a multidimensional anonymization methodology, this approach boosted scalability and reduced costs, but it required more research into evaluating privacy protection for huge datasets to be truly scalable.

Adaptive Utility-based Anonymization (AUA) was developed by J.J. Panackal and A.S. Pillai [8] as a smart first step toward preserving privacy. This method included numerous data sources and suppliers to personalize the QI characteristics for each user. A version of the model was anonymized so that it could provide the most benefits to users while still protecting their privacy (through k-anonymity), and so that as much data as possible could be made available to end users.

A technique for anonymizing data was created by Can Eyupoglu et al. [9], which maintains privacy while still making use of massive datasets. This method's scalability and practicality were evaluated by applying it to datasets of varying sizes. In this scenario, the data was disrupted with a chaotic function before being made public. When compared to other algorithms, this one has better results for the Kullback-Leibler divergence, F-measure, and precision.

To maintain confidentiality while maximizing data use in distributed and incremental datasets housed in the cloud,

X. Zhang et al. [10] developed a quasi-identifier index-based method. Quasi-identifiers were used to represent the anonymized data for the sake of efficiency. The method made use of computation-intensive programs to process big data, but it didn't factor in privacy-aware efficient scheduling for things like cloud-based data anonymization or the initiation of cloud-based privacy preservation.

With their polynomial-time approach, B. Srisungsittisunti and J. Natwichai [11] solved the problem of maintaining confidentiality. In this case, the computational complexity was drastically reduced while still guaranteeing the best possible outcome. When compared to other techniques for protecting individual privacy, this one is both successful and efficient. The method did not make an effort to investigate the complexity of requesting specific data. Operations like adding, removing, or querying the dataset still faced the huge hurdle of the concurrency control issue.

Using m-signature and fuzzy processing, J. Le et al. [12] developed a system that protects users' anonymity. Fuzzy processing was used in this strategy so that sensitive information may be kept secure while still being useful. This approach used a heuristic algorithm to perform update operations, which helped to minimize data loss when implementing the published information. When it came to revoking anonymity and linking specific data, the method also proved to be more secure. Updates to the privacy protection model can be carried out using the greedy heuristic method, the Niching evolutionary algorithm, or the Ching divination system.

To provide sufficient privacy protection while maximizing data value, Aldeen, Y.A.A.S., et al. [13] developed an anonymization technique that makes use of distributed and incremental datasets. In this methodology, the strategy is to implement privacy security in the most efficient way feasible, by making use of incremental and geographically separated information. The performance hit can be fine-tuned using the progressive anonymization method and aggregated anonymized datasets were used to keep data secure and private.

To ensure users of cloud services meet privacy standards, C. Liu et al. [14] developed a comprehensive privacy-preserving architecture. In this, your queries, your data, and even your storage locations are all kept secret thanks to the clever use of key-value query processing. Here, we develop a way to protect one's anonymity by employing a default approach based on commutative encryption. By striking a good compromise between security and performance, query speed was improved while keeping the cost of commutative encryption to a minimum. The study compared the practicability and efficiency of two distinct methods of implementation. This method did not work with more complicated queries like semi-join queries. It's also worth noting that the method didn't include any further cryptographic mechanisms to beef up the security of the framework.

These techniques have developed from a straightforward procedure for a single attribute to multi-attribute techniques, and they typically call for the creation of a dedicated transformed database for secondary usage. Since these techniques always add a noise factor to the mean zero, there is never any bias in calculating the mean. The two primary categories of approaches that depend on the data perturbation method are the probability distribution class and the fixed-data perturbation class. The original database is replaced with a different sample from the same distribution or with the distribution itself when using the security-control strategy in the first class. However, the second class that has been studied in the literature has only been created for categorical or numerical data. The methods for the second class are the main topic of this research.

In their most basic form, fixed-data perturbation methods involve introducing some noise term e to a secret attribute X to cause the disturbed attribute Y. With this technique, each attribute in a database with several attributes is altered independently of the others. $Y = X + e$, where e is taken from a probability distribution with the mean value of 0 and a known variance to the data, is the general description of this method. These techniques fall under the category of additive data perturbation (ADP). ADP techniques are not the only ones that can be utilized to generate aggregate statistics by preserving the people's privacy indicated in a database; Multiplicative Data Perturbation (MDP) is another option. In such an approach, the disturbed attribute Y is defined as $Y = Xe$ for a single confidential attribute X, in which e contains a mean value of 1.0 and a set variance. Bias is not present in calculating the mean because the average for $e = 1.0$. Every attribute is independently distracted from other characteristics when using the MDP method to alter multiple confidential attributes. Even though it is impossible to calculate the original values exactly for a single data record, it is a novel reconstruction approach to reliably estimate the distribution of the original data values. While it is inevitable that some information will be lost during the distribution reconstruction process.

Et Al. [66] Geometric data perturbation (GDP) methods have proven to be highly successful and Kanmaz has achieved a better privacy by using GDP along with random number generators using static data.

The sensitivity of a laplace function provides an upper constraint on the amount that its output needs to be changed for us to maintain our privacy, that is to say. The sensitivity of the laplace function acquires the magnitude due to which an individual's information varies the function f in the worst case. As a result, the introduced response's uncertainty hides the single individual's participation, which is represented by the $l1$ sensitivity of the laplace function. So, in our research, the laplace noise is considered for efficient data perturbations.

In geometric data perturbation methods, due to the better noise creation nature of Laplace noise, the privacy is preserved better. In the proposed research, considering the nature of laplace noise and geometric data perturbation, laplace noise is used after rotation and transformation techniques.

3. Methodology

3.1. Objectives:

- To provide a data publishing method with excellent privacy and attack resistance that is independent of the data set and applicable to various numerical properties.
- To minimize the time required in a distributed environment to train a model with IOT HealthCare data.

The implementation steps used in the proposed work:
Step 1: Select IOT Healthcare data.
Step 2: Identify the Sensitive attributes from the database.
Step 3: Apply Rotation on Sensitive attributes.
Step 4: Apply Scaling on Sensitive attributes
Step 5: Apply Translation on Sensitive attributes.
Step 6: Add Laplace Noise to the Sensitive attributes.
Step 7: Train the Model and evaluate the performance between the original dataset and perturbed dataset using LSTM model.
Seep 8: Configure ICA (Independent Component Analysis) to evaluate the attack resistance.

3.2. Geometric data transformations

Post removal of sensitive data, the published data may have additional information linked with other datasets for re-identifying the original data individuals. Hence, geometric data transformation is employed to preserve some sensitive attributes. Data transforming techniques are applied in sequence starting with the rotation of the data, scaling of the data, and translating of scaled data and finally adding an optimized noise to the data to increase the privacy of the data.
Representation of the steps explained in the formula below:

$$G(X) = R + T + S + \Delta \quad (1)$$

R-Rotation, T-Translation, S-Scaling and Δ- addition of noise

a) Rotation

Given challenges with the random rotation or projection matrix techniques, in terms of decreased data distance preciseness, the random rotation matrix is a good choice when it comes to maintaining data distance precisely, where data is rotated in an N-dimensional coordinate system on a chosen axis.

b) Translation and Scaling

The next step is scaling of the data and then the translation of data. In this method, features are Standardized by scaling the data to unit variance, and removal of the average is performed. Eliminating the mean and scaling to a unit variance are two methods for standardizing characteristics.

c) Laplace Noise

After completion of the above steps, an optimized noise (Laplace noise) is added to the data. The Laplace distribution is defined by two parameters: **location parameter μ** (mean) and **scale parameter b** (spread) and distribution peak is stronger and tails are fatter.
The general expression for the work on additive noise that was first publicized by Kim [12] is as follows.

$$Z = X + \varepsilon \quad (2)$$

Where Z indicates the transformed data point, X indicates the original data point, and ε indicates the random variable (noise) with a distribution $e \sim N(0, \sigma^2)$. After that, it is added to X. Finally, X is replaced with the Z for the published data set [13].

4. Results and Discussions

LSTM has been used to improve the accuracy. To evaluate the performance of the LSTM model below metrics have been used:

Precision: Precision is a measure of the accuracy of positive predictions. Higher precision indicates fewer false positives.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (3)$$

Recall: Recall measures the ability of a model to correctly identify positive instances out of all actual positive instances. Higher recall indicates fewer false negatives.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negatives} \quad (4)$$

Accuracy: Accuracy measures the overall correctness of predictions made by a model.

$$Accuracy = \frac{TP+FN}{TP+TN+FP+FN} \quad (5)$$

4.1. Dataset Description:

The following parameters of the patient are transferred through IOT sensors to the central database in the form of packets:

- Body Temperature,
- Glucometer,
- Mouth Airflow,
- Blood pressure,
- Pulse Oximeter,
- EMG,
- ECG Monitoring,
- Infusion Pump and
- GSR

Table 1. Dataset Description

Dataset	Description
IoT healthcare data	Total Instances:188694 Attributes:52

- **Dataset link:**

Click here to Access the dataset

frame.time_delta	frame.time_relative	frame.len	ip.src	ip.dst	tcp.srcport	tcp.dstport	tcp.flags	tcp.time_delta	tcp.len	tcp.ack	tcp.connection.fin	tcp.connection.rst
0.0	0.0	74	101612044	101612072	56808	1883	0x00000002	0.0	0	0	0.0	0.0
5.2E-5	5.2E-5	74	101612072	101612044	1883	56808	0x00000012	5.2E-5	0	1	0.0	0.0
8.0E-6	6.0E-5	74	101612044	101612072	56810	1883	0x00000002	0.0	0	0	0.0	0.0
1.2E-5	7.2E-5	74	101612072	101612044	1883	56810	0x00000012	1.2E-5	0	1	0.0	0.0
3.0E-6	7.5E-5	74	101612044	101612072	56812	1883	0x00000002	0.0	0	0	0.0	0.0
1.0E-5	8.5E-5	74	101612072	101612044	1883	56812	0x00000012	1.0E-5	0	1	0.0	0.0
4.0E-6	8.9E-5	74	101612044	101612072	56814	1883	0x00000002	0.0	0	0	0.0	0.0
9.0E-6	9.8E-5	74	101612072	101612044	1883	56814	0x00000012	9.0E-6	0	1	0.0	0.0
1.92E-4	2.9E-4	74	101612044	101612072	56816	1883	0x00000002	0.0	0	0	0.0	0.0
1.1E-5	3.01E-4	74	101612072	101612044	1883	56816	0x00000012	1.1E-5	0	1	0.0	0.0
4.0E-6	3.05E-4	74	101612044	101612072	56818	1883	0x00000002	0.0	0	0	0.0	0.0
1.0E-5	3.15E-4	74	101612072	101612044	1883	56818	0x00000012	1.0E-5	0	1	0.0	0.0
3.0E-6	3.18E-4	74	101612044	101612072	56820	1883	0x00000002	0.0	0	0	0.0	0.0
1.0E-5	3.28E-4	74	101612072	101612044	1883	56820	0x00000012	1.0E-5	0	1	0.0	0.0
3.0E-6	3.31E-4	74	101612044	101612072	56822	1883	0x00000002	0.0	0	0	0.0	0.0
1.0E-5	3.41E-4	74	101612072	101612044	1883	56822	0x00000012	1.0E-5	0	1	0.0	0.0
1.54E-4	4.95E-4	74	101612044	101612072	56824	1883	0x00000002	0.0	0	0	0.0	0.0
1.1E-5	5.06E-4	74	101612072	101612044	1883	56824	0x00000012	1.1E-5	0	1	0.0	0.0
1.94E-4	7.0E-4	66	101612044	101612072	56808	1883	0x00000010	6.48E-4	0	1	0.0	0.0
3.7E-5	7.37E-4	66	101612044	101612072	56810	1883	0x00000010	6.65E-4	0	1	0.0	0.0

only showing top 20 rows

Figure 2. Sample Dataset of IOT Healthcare data

4.2. Feature Selection and Extraction

Once the dataset is loaded, feature selection and extraction of data is done. This is mainly used to extract the required features and remove the recurring and unwanted features. After feature selection and extraction, the resultant data set is represented in Figure 3, (The number of columns has been reduced from 52 to 33) and the related correlation coefficient heat map (shows the relationship between variables and patterns or dependencies within the data.) is shown in Figure 4.

	frame.time_delta	frame.time_relative	frame.len	ip.src	ip.dst	tcp.srcport	tcp.dstport	tcp.time_delta	tcp.len	tcp.ac
0	0.000000	0.000000	74	101612044	101612072	56808	1883	0.000000	0	
1	0.000052	0.000052	74	101612072	101612044	1883	56808	0.000052	0	
2	0.000008	0.000060	74	101612044	101612072	56810	1883	0.000000	0	
3	0.000012	0.000072	74	101612072	101612044	1883	56810	0.000012	0	
4	0.000003	0.000075	74	101612044	101612072	56812	1883	0.000000	0	

5 rows × 33 columns

Figure 3. Dataset after feature selection and extraction

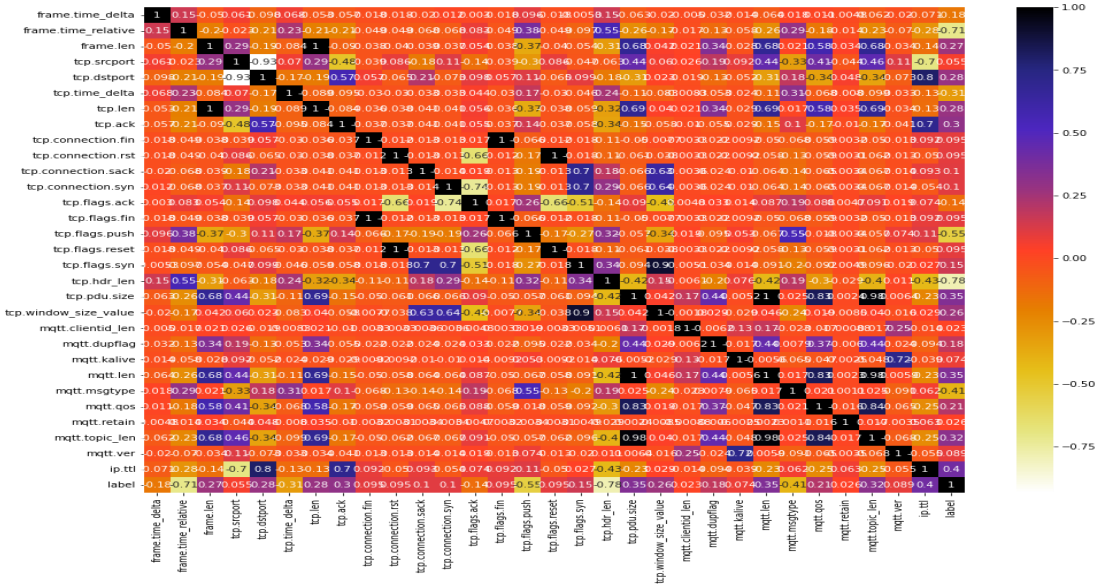

Figure 4. Correlation Coefficient Matrix (Heatmap View)

4.3. Geometric Data Perturbation

For the data obtained post-feature selection and extraction, the geometric data perturbations have to be performed on the dataset, which consists of rotation of the data, scaling of the data, and data translation, and adding of laplace noise.

4.4. LSTM

Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) capable of learning and remembering information over time, handling large volumes of data and ideal for big data applications, Model is trained iteratively over multiple epochs, with the parameters gradually adjusted to minimize the loss function and improve predictive performance.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 20, 20)            1760

 lstm_1 (LSTM)               (None, 20, 20)            3280

 lstm_2 (LSTM)               (None, 20)                3280

 dense (Dense)               (None, 1)                 21

=================================================================
Total params: 8,341
Trainable params: 8,341
Non-trainable params: 0
```

Figure 5. LSTM Model Summary

```
Epoch 1/100
1991/1991 [==============================] - 40s 17ms/step - loss: 0.1991 - val_loss: 0.1400
Epoch 2/100
1991/1991 [==============================] - 35s 17ms/step - loss: 0.1050 - val_loss: 0.0992
Epoch 3/100
1991/1991 [==============================] - 39s 20ms/step - loss: 0.0863 - val_loss: 0.0885
Epoch 4/100
1991/1991 [==============================] - 37s 18ms/step - loss: 0.0759 - val_loss: 0.0763
Epoch 5/100
1991/1991 [==============================] - 37s 18ms/step - loss: 0.0658 - val_loss: 0.0645
Epoch 6/100
1991/1991 [==============================] - 37s 18ms/step - loss: 0.0617 - val_loss: 0.0638
Epoch 7/100
1991/1991 [==============================] - 37s 18ms/step - loss: 0.0580 - val_loss: 0.0576
Epoch 8/100
1991/1991 [==============================] - 40s 20ms/step - loss: 0.0541 - val_loss: 0.0594
Epoch 9/100
1991/1991 [==============================] - 36s 18ms/step - loss: 0.0534 - val_loss: 0.0629
Epoch 10/100
```

Figure 6. Training of LSTM Model

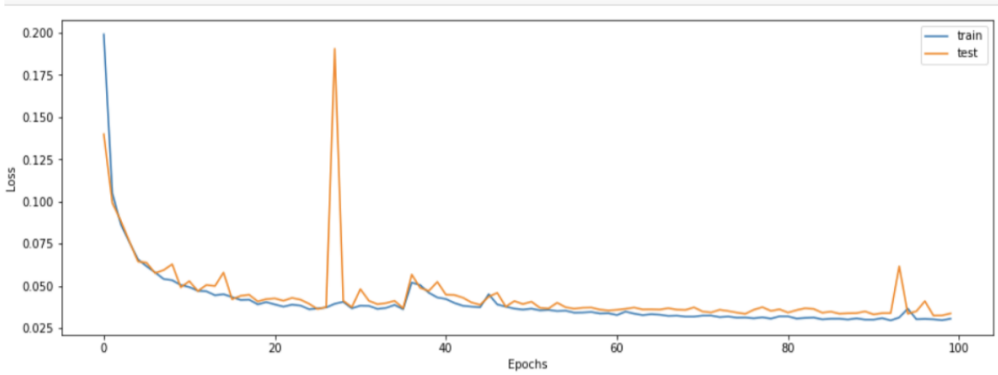

Figure 7. The learning curve of the LSTM Model

4.5. Performance Evaluation of LSTM vs Other Models

It's important to evaluate the LSTM model with methodologies like Decision Tree and Naive Bayes. The comparison of the model's performance with laplace noise vs LCG vs Gaussian noise is shown in Table 2.

4.6. Attack Resistance

Now there is always the probability of attacks on the dataset. Even after the application of novel techniques like laplace noise, attacks are probable. The attack resistance probability has to be evaluated and this is done using techniques like ICA, more details can be found in Table 3

Accuracy Comparison Results:

Table 2. Accuracy Comparison of LSTM with other Classifiers

Model	**Laplace Noise**	**Gaussian Noise**	**LCG Noise**
LSTM	0.8500	0.7866	0.7868
Decision Tree	0.7692	0.4960	0.7834
Naïve Bayes	0.8311	0.5621	0.5213

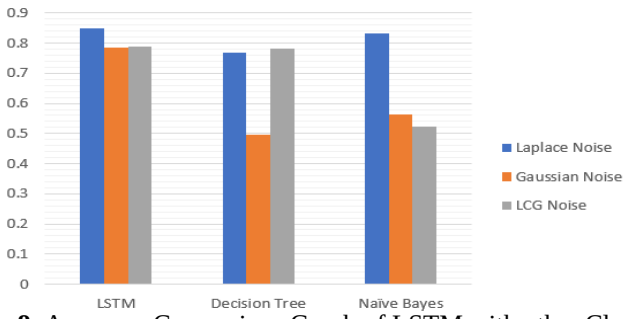

Figure 8. Accuracy Comparison Graph of LSTM with other Classifiers

Table 3. Attack Resistance Results

Data Set	Laplace	LCG	Gaussian
IoT healthcare data	0.9970	0.9888	0.9462

Table 4. Average execution time to generate the noise

Data Set	**Noise Algorithm**	**Execution Time**
IoT Healthcare	**Laplace**	2960
	LCG	3497
	Gaussian	3616

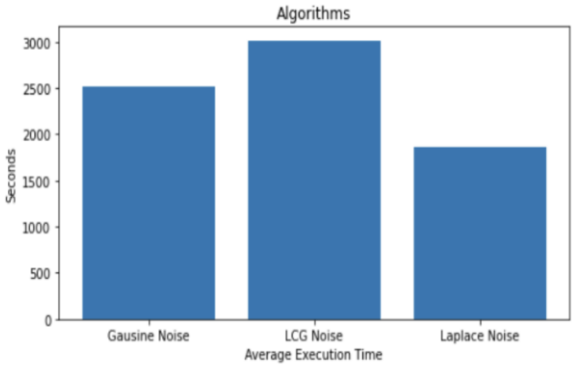

Figure 9. Average execution time to generate the noise.

5. Conclusion

We propose a new technique that integrates geometric data transformation with an optimized noise (laplace noise) for increasing data privacy and an LSTM model for improving Accuracy.

The suggested solution's performance assessment has been trained and tested with an IOT Healthcare data set (high in Velocity and Variety) using three different data privacy mechanisms, Laplace, LCG, and Gaussian noise methods along with GDTM.

The results have been compared with predictive Accuracy, defence against attacks (ICA), and speed of execution. A significant improvement in the Accuracy has been observed with the data privacy-preserving technique using Laplace noise and LSTM technique when compared with other Models and privacy-preserving techniques.

The average execution time to generate the laplace noise is faster when compared with other noise-generating techniques, which is a good factor to consider in real-time scenarios where speed matters in providing privacy.

The ICA results (Attack resistance) of the data have shown promising results with the Laplace mechanism when compared with LCG and Gaussian noise-generating techniques.

6. Future Work

There is a scope for improving the accuracy of the current privacy-preserving techniques by using advanced deep learning techniques like BERT, Attention Mechanisms, and gated Recurrent Units (GRUs) which are parallelizable architectures, making them more efficient unlike the LSTM networks which process data sequentially, especially for processing long sequences,

While numerical and textual data are important, image data is also equally important when it comes to preserving privacy, hence as a next step in my future work I would like to focus on Privacy-preserving of image data using the latest techniques in differential privacy or federated learning techniques.

References

[1] Chhinkaniwala I. and Garg S., DzPrivacy Preserving Data Mining Techniques: Challenges and issues, CSIT, 2011.

[2] L.Golab and M.T.Ozsu ,Data Stream Management issues, A Survey Technical Reportdz,´TT¸. 2018.

[3] Majid,M.Asger,Rashid Ali, DzPrivacy preserving Data Mining Techniques:Current Scenario and Future Prospectsdz, IEEE ´TT´, 2014.

[4] Malik, Majid Bashir, M Asger Ghazi, and Rashid Ali. “Privacy preserving data mining techniques: current scenario and future prospects.” Computer and Communication Technology (ICCCT), 2012 Third International Conference on. IEEE, 2012. 26–32.

[5] Aggarwal, Charu C. Data streams: models and algorithms. Vol. 31. Springer Science & Business Media, 2007.

[6] Chu, Fang. “Mining techniques for data streams and sequences.” Diss. University of California, Los Angeles, 2005

[7] Chang, Joong Hyuk and Won Suk Lee. “Finding recent frequent itemsets adaptively over online data streams.” Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2003. 487–492.

[8] Chaudhry, Nauman, Kevin Shaw, and Mahdi Abdelguerfi. Stream data management. Vol. 30. Springer Science & Business Media, 2006.

[9] Zhang, X., Yang, C., Nepal, S., Liu, C., Dou, W., and Chen, J., "A mapreduce based approach of scalable multidimensional anonymization for big data privacy preservation on cloud," in proceedings of Third International Conference on Cloud and Green Computing (CGC), pp. 105-112, September 2013.

[10] Panackal, J. J., and Pillai, A. S., "Adaptive utility-based anonymization model: Performance evaluation on big data sets," Procedia Computer Science., 50, pp.347-352, 2015.

[11] Eyupoglii, C., Aydin, M. A., Zahn., A. H., and Sertbas, A., "An Efficient Big Data Anonymization Algorithm Based on Chaos and Perturbation Techniques, "Entropy., 20(373), 2018.

[12] Zhang, X., Liu, C., Nepal, S., and Chen, J., "An efficient quasi-identifier index-based approach for privacy preservation over incremental data sets on cloud," Journal of Computer and System Sciences., 79(5), 542-555, 2013.

[13] Srisungsittisunti, B., and Natwichai, J., "An incremental privacy-preservation algorithm for the (k, e)-Anonymous

model," Computers & Electrical Engineering., 41, pp_126-141, 2015.
[14] Le, J., Zhang, D., Mu, N., Liao, X., and Yang, F., "Anonymous Privacy Preservation Based on m-Signature and Fuzzy Processing for Real-Time Data Release," IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2018.
[15] Aldeen, Y. A. A. S., Salleh, M., and Aljeroudi, Y., "An innovative privacy preserving technique for incremental datasets on cloud computing," Journal of biomedical informatics., 62, pp.107-116, 2016.
[16] Liu, C., Zhou, S., Hu, H., Tang, Y., Guan, J., and Ma, Y., "CPP: Towards comprehensive privacy preserving for query processing in information networks," Information Sciences., 467, pp. 296-311, 2018.
[17] S. Sangeetha, G. S. Sadasivam, "Privacy of Big Data: A Review," Springer International Publishing A. Dehghantanha, K. K. R. Choo (eds.), Book: Handbook of Big Data and IoT Security, India, pp. 5-23, 2019.
[18] Working party, "The working party on the protection of individuals with regard to the processing of personal data (0829/14/EN WP216)," Brussels, Belgium: Data Protection Working Party. http://statewatch.org/news/2014/apr/eu-art-29-dp-wp-216.pdf, 2014.
[19] T. Križan, M. Brakus, D. Vukelić, "In-Situ Anonymization of Big Data," IEEE Proceeding of the 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, pp. 292-298, 2015.
[20] S. Kavitha, S. Yamini, P. Raja Vadhana, "An Evaluation on Big Data Generalization Using k-Anonymity Algorithm on Cloud," IEEE Proceeding of the 9th International Conference on Intelligent Systems and Control (ISCO), Coimbatore, India, pp. 1-5, 2015.
[21] P. Jain, M. Gyanchandani, N. Khare, "Big data privacy: a technological perspective and review," Springer Journal of Big Data, Vol. 3, No. 25, 2016.
[22] Z. El Ouazzani, H. El Bakkali, "A new technique ensuring privacy in big data: K-anonymity without prior value of the threshold k," Elsevier Proceeding of The First International Conference on Intelligent Computing in Data Sciences, Morocco, pp. 52-59, 2018.
[23] U. P. Rao, B. B. Mehta, N. Kumar, "Scalable l-Diversity: An Extension to Scalable k-Anonymity for Privacy Preserving Big Data Publishing," IGI Global International Journal of Information Technology and Web Engineering (IJITWE), Vol. 14, No. 2, pp. 27-40, 2019.
[24] O. Hasan, B. Habegger, L. Brunie, N. Bennani, E. Damiani, "A Discussion of Privacy Challenges in User Profiling with Big Data Techniques: The EEXCESS Use Case," IEEE Proceeding of the International Congress on Big Data, CA, USA, pp. 25-30, 2013.
[25] K. M. PdShrivastva, M.A. Rizvi, S. Singh, "Big Data Privacy Based On Differential Privacy a Hope for Big Data," IEEE Proceeding of the 6th International Conference on Computational Intelligence and Communication Networks (CICN), Bhopal, India, pp. 776-781, 2014.
[26] L. Cui, Y. Qu, S. Yu, L. Gao, G. Xie, "A Trust-Grained Personalized Privacy-Preserving Scheme for Big Social Data," IEEE Proceeding of the International Conference on Communications (ICC), USA, pp. 1-6, 2018.
[27] M. Du, K. Wang, Z. Xia, Y. Zhang, "Differential Privacy Preserving of Training Model in Wireless Big Data with Edge Computing," IEEE Proceeding of The Transactions on Big Data, 2018.
[28] Data Masking, A Net 2000 Ltd. White Paper, "Data Masking: What You Need to Know, What You Really Need To Know Before You Begin", 2016.
[29] S. Arfaoui, A. Belmekki, A. Mezrioui, "Privacy Enhancement of Telecom Processes Interacting with Charging Data Records," Springer Proceeding of the Tenth International Conference on Soft Computing and Pattern Recognition (SoCPaR), pp. 268-277, 2018.
[30] K. Sharmila, A. C. S. Borgia, V.S. Sreeja, "A Comprehensive Study of Data Masking Techniques on cloud," International Journal of Pure and Applied Mathematics, Vol. 119, No. 15, pp. 3719-3728, 2018.
[31] V. Estivill-Castro and L. Brankovic. Data Swapping: Balancing Privacy Against Precision in Mining for Logic Rules. In Proc. of Data Warehousing and Knowledge Discovery DaWaK-99, pages 389–398, Florence, Italy, August 1999.
[32] R. Agrawal and R. Srikant. Privacy-Preserving Data Mining. In Proc. of the 2000 ACM SIGMOD International Conference on Management of Data, pages 439–450, Dallas, Texas, May 2000.
[33] Selva Rathnam S, Karthikeyan T. A survey on recent algorithms for privacy preserving data mining. International Journal of Computer Science and Information Technologies. 2015;6(2):1835-1840
[34] Patel A, Patel K. A hybrid approach in privacy preserving data mining. In: 2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA). Vol. 2. Ahmedabad, Gujarat, India: IEEE; 2016. p. 3
[35] M. Naga Lakshmi and K. Sandhya Rani, "A privacy preserving clustering method based on fuzzy approach and random rotation perturbation", Publications of Problems & Application in Engineering Research-Paper, Vol. 04, Issue No. 1, pp. 174-177, 2013.
[36] Mary AG. Fuzzy–based random perturbation for real world medical datasets. International Journal of Telemedicine and Clinical Practices. 2015;1(2):111-124. DOI: 10.1504/IJTMCP.2015.069749
[37] M. Naga Lakshmi, K Sandhya Rani," Privacy preserving hybrid data transformation based on SVD"," International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 8, 2013, 2278-1021
[38] Jalla HR, Girija PN. An efficient algorithm for privacy preserving data mining using hybrid transformation. International Journal of Data Mining & Knowledge Management Process. 2014;4(4):45-53. DOI: 10.5121/ijdkp.2014.4404

[39] Manikandan G, Sairam N, Saranya C, Jayashree S. A hybrid privacy preserving approach in data mining. Middle- East Journal of Scientific Research. 2013;15(4):581-585. DOI: 10.5829/idosi.mejsr.2013.15.4.1.991
[40] Saranya C, Manikandan G. Study on normalization techniques for privacy preserving data mining. International Journal of Engineering and Technology (IJET). 2013;5(3):2701-2704
[41] Geetha Mary AN, Iyenger NSC. Non-additive random data perturbation for real world data. Procedia Technology. 2012;4:350-354. DOI: 10.1016/j.protcy.2012.05.053
[42] Aggarwal CC, Yu PS. A condensation approach to privacy preserving data mining. In: Proceedings of International Conference on Extending Database Technology (EDBT). Vol. 2992. Heraklion, Crete, Greece: Springer; 2004. pp. 183-199
[43] Liu K, Kargupta H, Ryan J. Random projection-based multiplicative data perturbation for privacy preserving distributed data mining. IEEE Transactions on Knowledge and Data Engineering (TKDE). 2006;18(1):92-106
[44] Chen K, Liu L. "A Random Rotation Perturbation Based Approach to Privacy Preserving Data Classification", CC-Technical Report GIT-CC-05-12. USA: Georgia Institute of Technology; 2005
[45] Lui K, Giannella C, Kargupta H. An Attacker's view of distance preserving maps for privacy preserving data mining. In: Proceedings of the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases(Pkdd'06). Berlin, Heidelberg: Springer-Verlag; 2006
[46] Xu H, Guo S, Chen K. Building confidential and efficient query services in the clod with RASP data perturbation. IEEE Transactions on Knowledge and Data Engineering. 2014;26(2):322-335
[47] Oliveira SR, Zaiane OR. Privacy preserving clustering by data transformation. Journal of Information and Data Management (JIDM). 2010;1(1):37–51
[48] Guo S, Wu X. Deriving private information from arbitrarily projected data. In: Proceedings of the 11th European conference on principles and practice of knowledge Discovery in databases (PKDD07). Warsaw, Poland. 2007
[49] Balasubramaniam S, Kavitha V. A survey on data retrieval techniques in cloud computing. Journal of Convergence Information Technology. 2013;8(16):15-24
[50] Liu J, Yifeng XU. Privacy preserving clustering by random response method of geometric transformation. Harbin, Heilong Jiang, China: IEEE. 2010:181-188. DOI: 10.1109/ICICSE.2009.31
[51] Balasubramaniam S, Kavitha V. Geometric data perturbation-based personal health record transactions in cloud computing. The Scientific World Journal. 2015; 2015:927867, 1-927869. DOI: 10.1155/2015/927867
[52] Chen K, Lui L. Geometric Data Perturbation for Privacy Preserving Outsourced Data Mining. London: Springer-Verlag Limited; 2010
[53] Hyvarinen AK, Oja E. Independent Component Analysis. New York/Chichester/Weinheim/Brisbane/Singapore/Toronto: Wiley-Interscience; 2001
[54] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation
[55] Graves, A., Mohamed, A. R., & Hinton, G. (2013). Speech Recognition with Deep Recurrent Neural Networks. In Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on (pp. 6645-6649). IEEE., 9(8), 1735-1780.
[56] Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks. In Advances in Neural Information Processing Systems 27 (pp. 3104-3112).
[57] Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A Search Space Odyssey. IEEE Transactions on Neural Networks and Learning Systems, 28(10), 2222-2232.
[58] Aggarwal, C. C., & Yu, P. S. (2008). A General Survey of Privacy-Preserving Data Mining Models and Algorithms. Privacy-Preserving Data Mining: Models and Algorithms, 11-52.
[59] Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E. S., Spicer, K., ... & de Wolf, P. P. (2012). Statistical Disclosure Control.
[60] Dwork, C. (2006). Differential Privacy. In Automata, Languages and Programming (pp. 1-12). Springer, Berlin, Heidelberg.
[61] Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep Learning with Differential Privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (pp. 308-318).
[62] Dwork, C., & Roth, A. (2014). The Algorithmic Foundations of Differential Privacy. Foundations and Trends in Theoretical Computer Science, 9(3-4), 211-407.
[63] Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating Noise to Sensitivity in Private Data Analysis. In Theory of Cryptography Conference (pp. 265-284). Springer, Berlin, Heidelberg.
[64] Dwork, C., & Roth, A. (2014). The Algorithmic Foundations of Differential Privacy. Foundations and Trends in Theoretical Computer Science, 9(3-4), 211-407.
[65] Dwork, C., & Rothblum, G. N. (2016). Concentrated Differential Privacy. In Proceedings of the 2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS) (pp. 66-75). IEEE.
[66] Kanmaz, M., Aydın, M. A., & Sertbaş, A. (2021). A New Geometric Data Perturbation Method for Data Anonymization Based on Random Number Generators. Journal of Web Engineering, 20(6), 1947–1970. https://doi.org/10.13052/jwe1540-9589.20613