Journal of Critical Reviews

ISSN- 2394-5125 Vol 7, Issue 14, 2020

A NOVEL PRIVACY PRESERVING BASED ENSEMBLE DEEP LEARNING FRAMEWORK FOR LARGE DIMENSIONAL MEDICAL DATASETS

[1]Ch.Nanda Krishna, [2]Dr. K.F.Bharati,

[1]Research scholar, Department of computer science and engineering, JNTUA,Anantapuramu.
[2]Assistant professor, Department of computer science and engineering, JNTUACEA, anantapuramu.

Received: 06.04.2020 ***Revised: 08.05.2020*** ***Accepted: 03.06.2020***

Abstract:
With the rapid growth of data size, computing power and high dimensionality, it is essential to implement a novel privacy preserving model in deep learning framework. However, most of the traditional machine learning applications are integrated in deep learning framework on large databases, it is essential to secure the large sensitive patterns that are generated by the machine learning approaches before uploading to the cloud storage. As a result, how to design and implement a novel privacy preserving deep learning model (PPDLM) over high dimensional cloud data becomes a challenging task. Traditional privacy preserving deep learning frameworks are depends on data transformation approaches rather than the cryptographic approach due to high computational memory and time in cloud computing. In the real time multi-user applications, multiple datasets are distributed across the multiple users for privacy preserving. As the data size of multi-user applications increases, traditional PPDLM models require high computation memory and time for preserving the machine learning patterns. To overcome these problems, a novel data partitioning based privacy Preservation deep learning model is implemented on high dimensional datasets. Experimental results proved that the present system has high computational accuracy with privacy in the patterns compared to the existing models.

Keywords: high dimensionality, PPDLM models, cryptographic approach

DOI: http://dx.doi.org/10.31838/jcr.07.14.01

INTRODUCTION

Deep learning is a multi-layered data processing network that consists of multiple levels of abstraction to train the data for pattern analysis[1]. This network use a non-linear transformation approach to transform and learn the data in each level. Recently, a large number of composite functions have been used in the deep learning framework for pattern analysis. The deep learning system has recently been applied in almost all fields of research, technology, medical science , education, the business and governmental organizations, etc. The organizations started to seek a deeper, less computational-time and memory privacy from large databases[2] for future prospects. Machine learning is the state-of-the-art technology from the artificial intelligence industry that extract from vast datasets secret information, or laws. In master learning associations are used for privacy security applications in a number of applications, grouping approaches, clustering approaches and possible predictive approaches.

The information about the individuals or company related data are collected and maintained at remote cloud server which contains sensitive data about the individuals and hence cannot be disclosed to third party authorities. Moreover, the patterns generated on the distributed data using the deep learning models can also lead to disclosure of sensitive information without privacy preserving approaches[3]. Hence the data owners publish their data only when there is a guarantee for the privacy of data. Most of the traditional deep learning based privacy preserving models are not appropriate for selecting the essential privacy patterns due to high computational time and memory. Also, most of the traditional deep learning based privacy preserving models are mainly focus on the limited data from a single source. A deep learning platform has recently been used to enhance the efficiency of data processing approaches including the extraction of functional data, classification, medical imaging segmentation. The deep neural network (DNN) has the ability to deep-layer patterns. DN Network (DNN) and Convolution (CNN) are among the most commonly used deep-neural network architectures for data processing[4].

Data publication masks the data using a transformation or cryptographic approach before the data is shared with third parties. At the first step it is unknown how the data are to be used. Therefore, data preservation technology converts the data into a calculation-friendly form. For any usage not explicitly in data mining, the published data may be used. Data randomization is a transformation technique that would adjust the value of the original operation, such as scaling , rotation or noise addition .. During the randomisation process the mapping or correlation between the data is lost completely. Cryptographic solutions use specific encryption algorithms to prevent intruders from being stored. However, once the data has been decrypted, the data becomes easy and a half-honest opponent will lead to attacks[5].

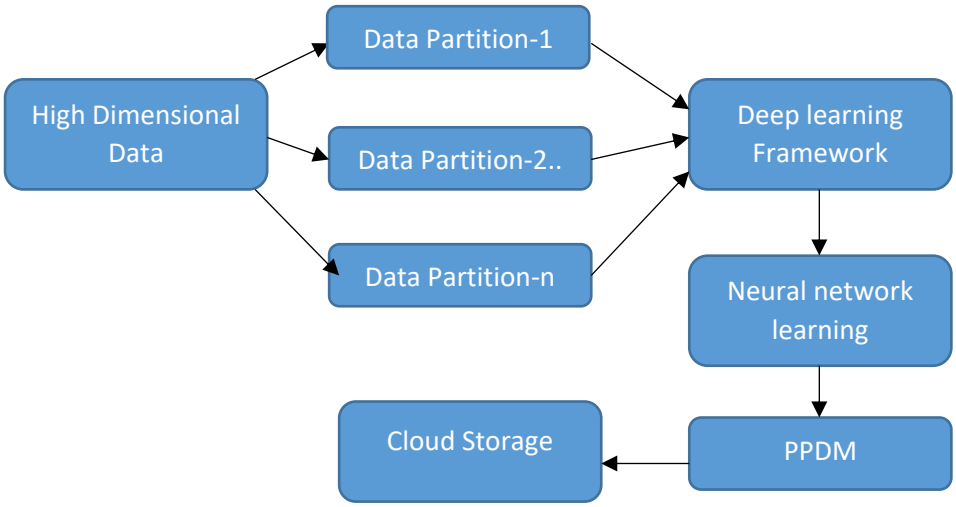


Figure 1: Basic Privacy Preserving deep learning framework on high dimensional data

Figure 1, describes the basic approach of the traditional privacy preserving deep learning framework on high dimensional datasets. In this framework, the data size is restricted due to high computational memory and time. As shown in the figure, each input dataset is partitioned to train the data using the deep learning framework. Neural network model is used to learn the partition for data prediction. Each predicted value is privacy preserved using the differential privacy preserving approach. Finally, predicted patterns are stored in the cloud server[6].

Data partitioning, there are two scenarios that require using of cluster analysis in a distributed way. In the first, the volume of data that is to be analysed is fairly great. Therefore, this requires a huge amount of computational effort—so much so, sometimes, it is not feasible to complete this computation. In such a case, a better alternative is to split the data and cluster it in a distributed manner and, finally, unify the distributed results. In centralized database, data will be located and maintained at single place where as in distributed database, data may be distributed vertically or horizontally to various sources. When the database is centralized, all the data is stored in one place. This type of database is completely different from the distributed database. One of the issues the centralized database faces is that as the entire data resides at one central location, there can be problems with bottle-necks occurring at key points where the data is released or assimilated[7]. As a result, when looking for the availability of data, the efficiency with which it is retrieved is not as strong as in the distributed database system.

The main contribution in this paper include,

1) Multi-user data partitioning is performed to improve the runtime of the data preprocessing.
2) Multi-layered data pre-processing approach is implemented to remove the noise or sparse values in the data partition.
3) Privacy preserving based ensemble deep learning framework is designed and implemented on the filtered datasets.

The rest of the paper is organized as follows. Section 2, describes the related works of the privacy preserving models and its limitations. Section 3, describes the proposed solution to the privacy preserving based deep learning framework on high dimensional data. Section 4, describes the experimental results and analysis. Finally, we conclude the paper in section 5.

RELATED WORKS

In unsupervised self-organizing learning with support vector rankings as stated in[8], the imbalanced data problem is relaxed. The model adopted by support vector machines in this method selects variables for dealing with this problem. Also known as the Emergent Self-Organizing Diagram, ESOM is used to cluster the ranker features to provide unsupervised classification of clusters. A Kolomogrov-Smirnov statistics based on decision tree system(K-S tree)[9] is the latest approach by which complex problems are divided into several simpler sub-problems, in which case unequal distribution is less challenging. The classification algorithm is implemented in a multi-class dataset for overcoming the said problem. Compared to features of strength, the Gaussian texture features algorithm shows better results. An effective weighting technique can differentiate between database scans from different domains, based on classifiers. This methodology will definitely improve the performance of the conventional approaches to classification.

Preserving privacy in data is achieved by hiding the original features of the data to the end user. The data mining technologies work on original data to extract the knowledge from the collected huge data sets. The data sets are sensitive, containing information about the privacy features of individuals or any other confidential operations. Such data if directly processed by the data miners; certainly will reveal the sensitive information and thus becomes a cause for privacy breach. A number of measures for anonymity are available in the data mining industry, which are usually defined as data anonymization methods as shown in fig1. The algorithms for preserving privacy are differed into two groups, there are distribution frameworks and methods that add random noise to the original source of data.

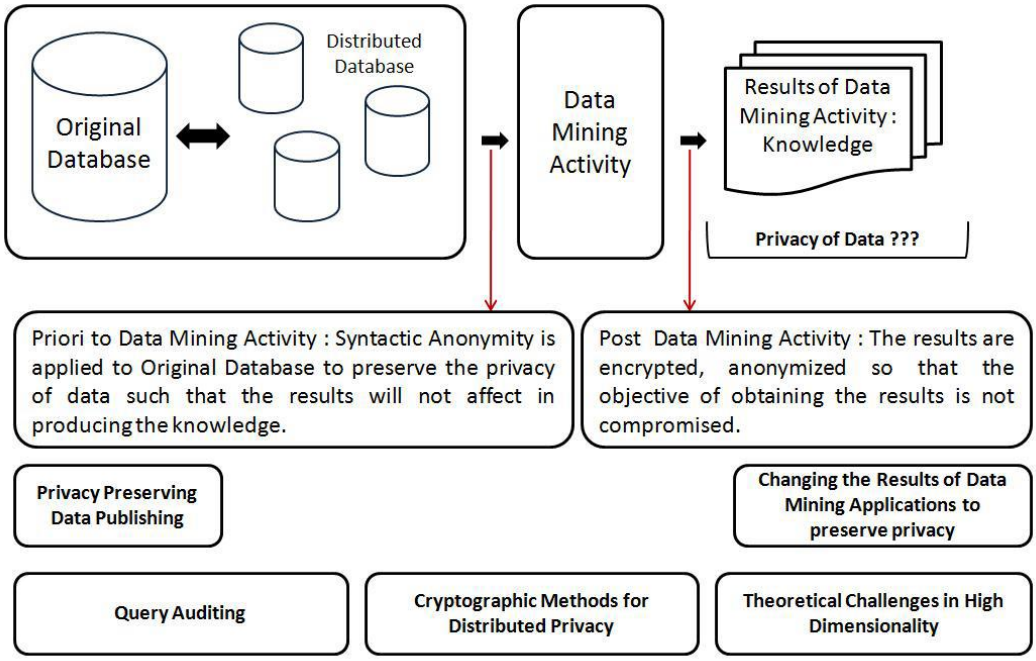


Figure 1: Traditional cryptographic based privacy preserving model

The Privacy Preserving Data Mining (PPDM) problem in this traditional work concentrates on two important aspects. The first facet of the research: Assuring privacy of database based on the trust levels of the analysts and with respect to the key attributes for their data mining queries. The second facet of the research is to assess the sensitivity level of the information that is disseminated from the database based on the analysts‟ queries. The issue of utility-based privacy controlling data mining was reviewed in [10]. The overall plan in [11] is to identify and reduce the impact of structuring by individually publishing tables which contain the utility based attributes, but it creates issues the basic need of preserving privacy. The details performed on the real tables and the marginal tables need not be same. It is now identified that this approach can improve managing and preserving the data set without negotiating on the privacy.

Yuan et.al, proposed an advanced statistical and artificial neural network-based analysis in order to describe complexity and heterogeneity in preeclampsia [12]. Preeclampsia is considered as the most common disease that happens at the time of pregnancy. This disease can increase the blood pressure and also affects liver and kidneys. If the disease is not treated in its early stage, then it can be harmful for both mother and the baby. The above proposed method can analyse microarray data sets in order to solve the complexity issues and handle the heterogeneity property. In this research paper, a new statistical multiple comparison based technique is introduced which is beneficial in order to detect the symptoms of preeclampsia. The above features are helpful during the construction of an artificial neural network based machine learning scheme. Again, this machine learning approach is implemented in order to delineate features. The functional analysis of the features can be carried out with the help of randomisation technique. Further research efforts can be performed in order to extend the above presented technique.

ADVANCED PRIVACY PRESERVING BASED DEEP LEARNING FRAMEWORK

In the proposed ensemble deep learning based privacy preserving framework, a novel data pre-processing and attribute partitioning methods are proposed to minimize the noise and data processing on high dimensional datasets.

In this model, each multi-user dataset is pre-processing using the novel data transformation method. This transformation method is used to remove the missing values and sparsity values on large datasets. After the data pre-processing step, a novel ensemble deep learning model is applied on the filtered data for privacy preserving. Proposed ensemble deep learning model consists of three classification approaches such as partitioning decision tree, non-linear SVM and deep multi-functioned neural network approaches as shown in fig 2.

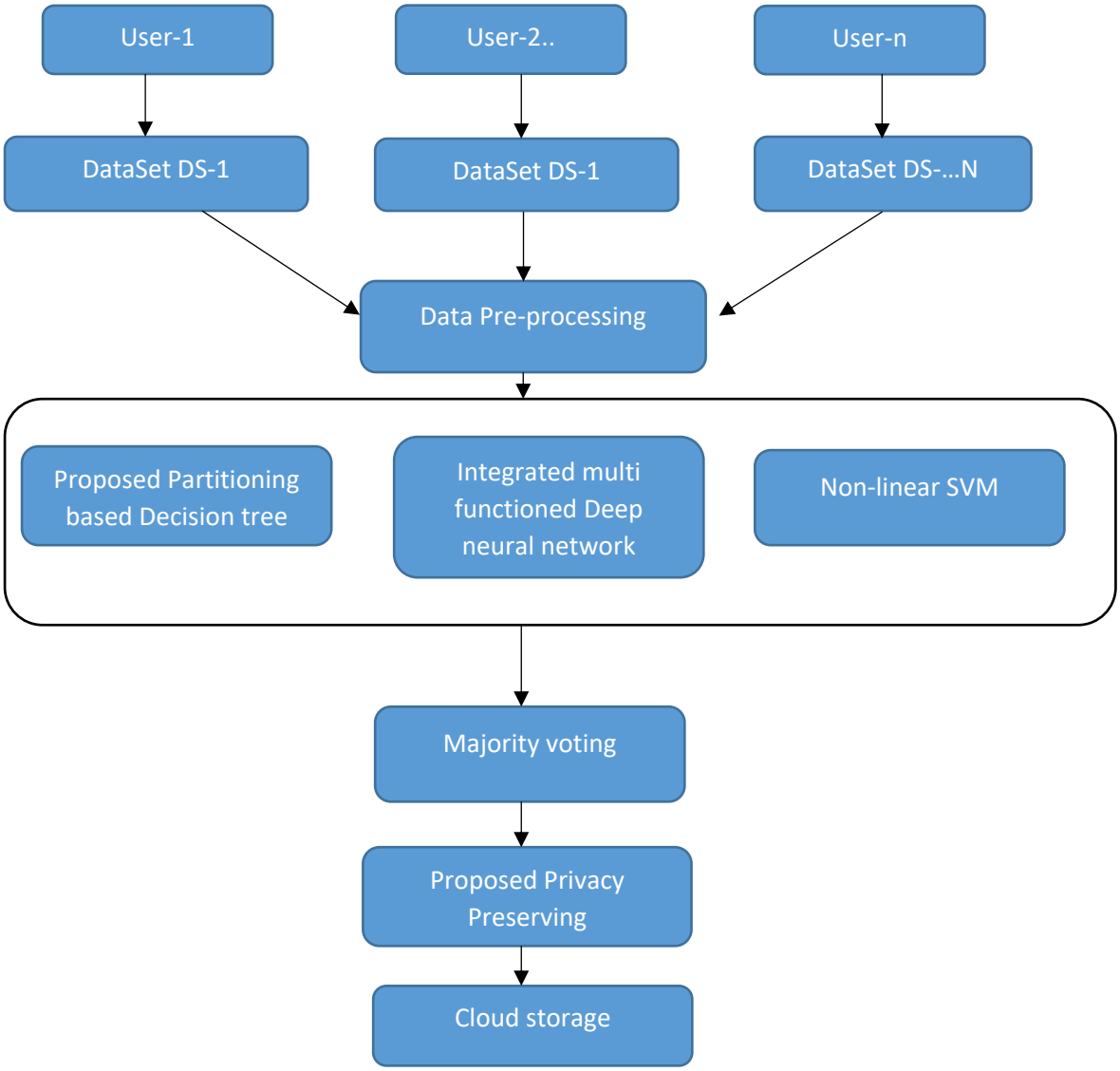


Figure 2: Proposed Ensemble deep learning framework for Privacy preserving

Here, an optimized integrity based homomorphic privacy preserving model is developed for the deep learning patterns. Finally, these sensitive privacy patterns are published in the remote cloud storage.

Algorithm 1: Multi source data pre-processing

Input : Multi source datasets MD={D-1,D-2...D-n}, Attribute: $A\tau$,Max attribute value M_x, Minimum attribute value M_n, frequency of the maximum attribute value that contains class c Max_c, frequency of the minimum attribute value that contains class c Min_c,

1: Read input datasets MD
2. For each dataset D[i]
3. Do
4. For each record I[r]
5. Do
6. For each attribute in I[r]
7. Do
8. If($A\tau$ [I]==Continuous && $A\tau$ [I]== ϕ)
9. then
10. Replace $A\tau$ [I] using the eq .(1)
11. $$A\tau[I] = \frac{A\tau[I] - (M_x(A\tau) + M_n(A\tau))/2}{Max_c(A\tau) - Min_c(A\tau)} \quad \text{--(1)}$$
12. End if
13. If($A\tau$ [I]==Categorical && A[I]== ϕ)
14. Then

15. Replace $A\tau$ [I] using the eq.(2)

16.
$$A\tau[I] = \frac{\sum_{i=1,m=1} F(A\tau[i]/c_m) - F(c_m)}{M_X \, Prob(A\tau[i]/c_m)} \quad \text{-----(2)}$$
$$i = 1..n; m = 1..k(classes)$$

17 Done
18 Done
19 done

In the algorithm 1, each data instance from the distributed data source is pre-processed using the min-max measure and max probabilistic measure. In each data source, each instance is pre-processed by using numerical or nominal type of attribute. If the attribute type is continuous, then the equation (1), is used to fill the missing value of the numerical attributes. Similarly, in case of nominal or categorical attributes, conditional probability of the attribute is used to fill the missing value.

Algorithm 2: Ensemble Deep learning framework
Input : Pre-processed dataset PD-1,PD-2..PD-N, Fast Random Decision tree with Node N, Tree count |N|, Attributes list A.
Procedure:

Fast Random forest

1. For each attribute list in $A\tau$
2. Do
3. If(P-D[i]==Null)
4. Then
5. Make leaf node D.
6. End if
7. End for
8. For each node n in |N|
9. Do
10. If (n[i]==|PD|) // nodes count is equal to total number of instances
11. Create leaf node n[i].
12. End for
13. For each attribute $A\tau$
14. Do
15. Partition $A\tau$ using the different class labels using the following measure.
16. $$\text{RandomForest Partitioning Measure=RPM[Att,m]=} \frac{\text{-PD[i].log}(\sum_{i=1}^{|PD|} PD[i])*Prob(PD[i]/C_m)}{PD[i]^3.\sqrt{Chisquare(PD[])}}$$
17. End for
18. End for
19. Choose the node with class m as the best split attribute in partitioning list.
20. Create root node N[0].
21. Repeat the process until all the nodes in the Tree T.
22. For each pattern in the Tree T
23. Do
24. Apply Advanced Integrity homomorphic encryption in the antecedent rule or consequent rule based on the user's choice.
25. Patterns P[A[i]]=AIH(pattern,ruletype);
26. End for
27. End for

Non-linear SVM

Apply SVM multi-class optimization models as

$$\min_{W_k, a_k} \frac{1}{2}||W_k||_1^2 + \alpha \sum_{i=1}^{n} ker < x, y > + \tau_m$$

$$s.t$$

$$W_k^T D_i + b_k \geq 1 - \tau_m, if \; y_i = k$$

$$W_k^T D_i + b_k \leq -1 + \tau_m, if \; y_i \neq k$$

$$\tau_m > 0; m = 1...classes$$

Here kernel function ker(x,y) defines the x input values that are mapped to y dimensional space as:

Ker<x,y>= $e^{-||y||\sum||x||^2}.max\{||x||,||y||\}$ if x==y

= $e^{-||y||\sum||x||^2}$ if x<y

$= e^{-||y||\sum||x||^2}$ if x>y

Multi-functioned Deep layered Network

Step 1: Assign feature weighted using the maximized weights using the (1) ,(2) and (3) to each feature.

T-statistic weighting measure is used to find the variation in the features using the standard deviation of the class labels. Basically, it is the ratio of difference of the means of the class labels to the maximized standard deviation.

$$W1 = \frac{\mu_P - \mu_N}{\sqrt{\min\{\sigma_P^2, \sigma_N^2\}}} \quad \text{-----(1)}$$

where μ_P is the mean of the positive class samples

μ_N is the mean of the negative class samples.

It is the ratio of difference of the means of the class labels to the sum of the standard deviation of the positive and negative disease classes. Here, the patterns with highest signal to noise ratio measure is selected as highest weighting measure for data classification.

$$W2 = HSNR = \frac{|\mu_i - \mu_j|}{2(\sigma_P * \sigma_N)} \quad \text{-----(2)}$$

where μ_P and σ_P are the mean and standard deviation of the positive class samples

μ_N and σ_N are the mean and standard deviation of the negative class samples.

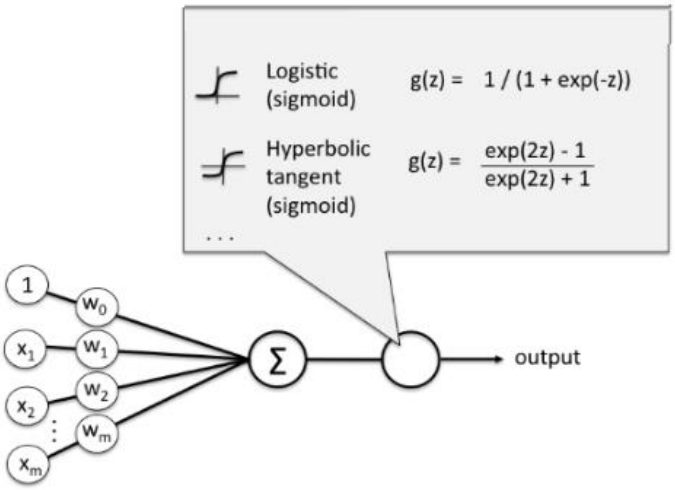
It is the maximization of the correlation between the features, hybrid t-test and hybrid SNR ratio. This ranking measure is used to select the optimal binary class features in each cluster.

$$W3 = MCTSNR = Max\{Roughset(Features : F), W1, W2\} \quad \text{-----(3)}$$

Weights W[]=Average{W1,W2,W3}

Step 2: Defining the input, hidden and output layers to each mapper for parallel processing.

Step 3: To each hidden layer apply the logistic activation function for weights and error rate optimization.



Step 4: Classify data using the deep neural network framework until error rate and weights are converged.

To each pattern in the decision tree construction, rule type is considered as either left side or right side of the pattern for privacy preserving. The advanced integrity based homomorphic encryption scheme used to preserve the left side or right side of the pattern is described below:

AHomo(Pattern P,ruletype):

Step 1: Converting pattern P into byte array PB[].

Step 2: Parameter initialization

In this phase, public key and master key are generated using the bilinear pairing and cyclic group elements.

$$\textbf{PublicKey}\ Pub_k^V = \{g_p, g \in G \wedge H; e(g, g^{\alpha})\}$$

$$Masterkey = \{g_{\alpha}, \beta \in G \wedge H\}$$

Choose two random prime numbers k1,k2.

n=k1*k2;

s=n*n;

Choose a random noise $r_n \in (0,1)$

$$\lambda = \frac{k1.k2}{(n^{(k1)} \bmod(r_n))^{(k2)} \bmod(r_n)}$$

$$r = \lambda^{\gcd(k1.k2, r_n)}$$

Step 3: Encryption

$$CB[]=E(PB[])= r_n^{PB[i]} \bmod(s).r^n \bmod(s).\bmod(s)$$

Step 4: Decryption D(CB[])

$$h_1 = r_n^{\lambda} \bmod(s) - (\frac{\lambda}{n})^{-1} \bmod(n)$$

$$PB[]= CB[i]^{\lambda} \bmod(s) - \frac{\lambda}{n}.h_1 \bmod(n)$$

Third party Privacy Preserving Data Change Detection

Input: Ensemble classifier patterns, Pre-computed pattern integrity PI, ψ m_1, $g_{\theta 2}$ $\theta 1, \theta 2, \lambda_1$ are the randomized cyclic group elements. Pub_k^V is the public key which is generated in the initialization step. H(Data) is the computed whirlpool hash value, M_D is the multi-user data.

Let $T_\alpha, T_\beta, T_\gamma$ represents the cyclic group elements with multiplicative operation. g_{m1}, g_{m2} are the cyclic group generators with m1 and m2 $\in T_{m1}, T_{m2}$.

ψ be the transformation cyclic group function defined from $G_{m2} \to G_{m1}$ $\ni$ $\psi(g_{m2}) = g_{m1}$.

Let 'e' be the bilinear map of the cyclic group as $e \in G_{m1} \times G_{m2} -> G_{m3}$.

Cloud server (CS) selects a randomized security parameter $k \in R(q,*)$.

$$CloudCParams = \{M\_D, m1, \theta 1 \in G_1, m2, \theta 2 \in G_2, m3, \theta 3 \in G_3, g_{\theta 1}, g_{\theta 2}, g_{\theta 3}, e, Pu_k^{MC} = \{g_{m1}^k\}, \lambda_1 = MH(M\_D)$$

$$\lambda_2 = g_{\theta 1}^{\frac{1}{(x_1+\lambda_1)\psi}}, x_1 \in g_{\theta 1}, MH\{I_{256}((M\_D \| Data)), I_{512}((M\_D \| Data), I_{1024}((M\_D \| Data), I_{4096}((M\_D \| Data), \}\}$$

Here, each cloud parameters contains various security related parameters such as M_D is the multi-user dataset, Pu_k^{MC} is the user public key, MH is the set of multi-user hash integrity computed values with different key size.

Output: Data change Detection using integrity computation and signature.

1. For each multi-user MC
2. Do
3. Find the patterns that are related to MC.
4. PMC[]=getPatterns[MC[i]];
5. For each pattern PMC[]
6. Do
7. Compute current Integrity using the ensemble learner patterns as IC(PMC[j])
8. If(IP(PMC[j])==IC(PMC[j])
   Then

$$e(\lambda_2^{\psi}, Pub_k^V, g_{\theta 1}^{H(M\_ID)}) = e(g_{\theta 2}^{\frac{\psi}{\psi(m_1+\lambda_1)}}, g_{\theta 1}^{m_1}.g_{\theta 1}^{\lambda_1})$$

$$= e(g_{\theta 2}^{\frac{1}{(m_1+\lambda_1)}}, g_{\theta 1}^{m_1+\lambda_1})$$

$$= e(g_{\theta 2}, g_{\theta 1})^{\frac{1}{(m_1+\lambda_1)}(m_1+\lambda_1)}$$

$$= e(g_{\theta 2}, g_{\theta 1})^1$$

$$S_{MC} = e(g_{\theta 2}, g_{\theta 1})$$

9. if($S_{MC} \equiv e(g_{\theta 2}, g_{\theta 1})_{MC}$)
   then
   Patterns are not changed by third party authority.
   Else
   Patterns are changed by third party authority.
   Done

Done

EXPERIMENTAL RESULTS

Experimental results are simulated in AMAZON AWS cloud environment. Here, distributed medical datasets are taken as input from different sources and then privacy preserving model is applied on the centralized database. In the proposed model, a novel integrity based homomorphic encryption is used to preserve the privacy for the ensemble deep learning model. Experimental results proved that the present model is better than the traditional methods in terms of privacy and accuracy are concerned.

Table 1: Comparative analysis of present homomorphic model to the existing models in terms of privacy preserving on one partitioning attribute.(Max depth=1)

Instances	AES	ECC	Homomorphic Encryption	Adaptive IntegerityHomomorphic
#50	0.919	0.942	0.951	0.976
#100	0.922	0.938	0.936	0.98
#150	0.924	0.929	0.932	0.972
#200	0.924	0.942	0.946	0.975
#250	0.917	0.924	0.941	0.975
#300	0.924	0.925	0.954	0.977
#350	0.922	0.932	0.952	0.984
#400	0.915	0.924	0.924	0.975
#450	0.914	0.931	0.928	0.984
#500	0.918	0.925	0.951	0.98
#550	0.915	0.936	0.928	0.973
#600	0.912	0.928	0.929	0.977
#650	0.919	0.944	0.951	0.979
#700	0.925	0.928	0.935	0.98
#750	0.923	0.929	0.946	0.977
#800	0.918	0.94	0.939	0.982
#850	0.919	0.926	0.941	0.976
#900	0.917	0.942	0.953	0.978
#950	0.92	0.924	0.936	0.975
#1000	0.914	0.928	0.951	0.975

Table 1, explains the efficiency of the model proposed to conventional models for privacy protection. The model preserving privacy performs better in the present system than the traditional models on one partitioning attribute.(Max depth=1).

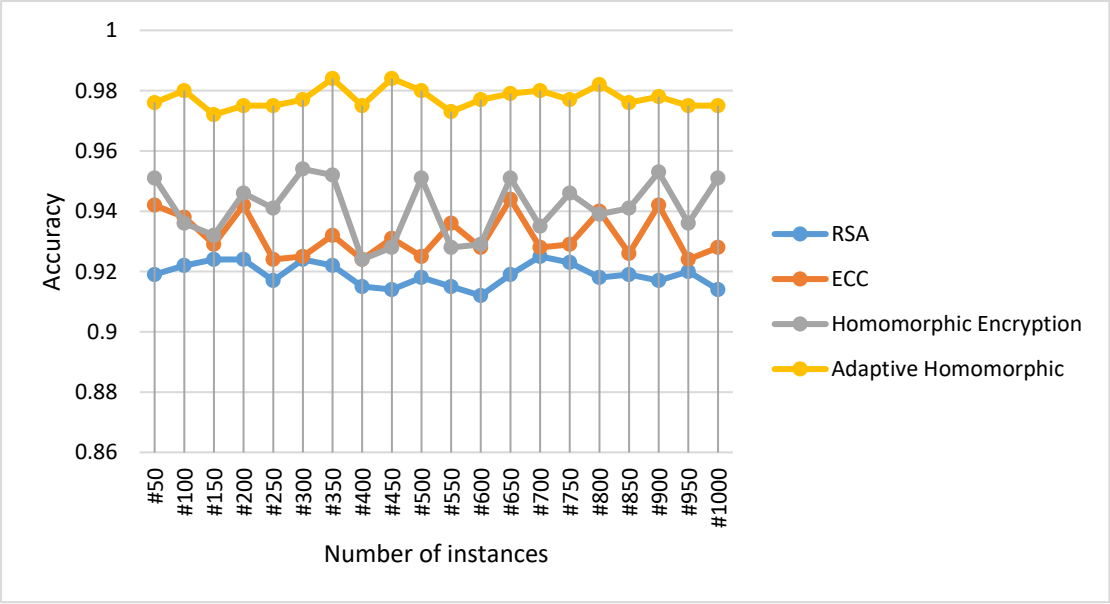


Figure 3: Comparative analysis of present homomorphic model to the existing models in terms of privacy preserving on one partitioning attribute.(Max depth=1).

Figure 3, explains the output of the proposed model to conventional pattern preservation models for privacy. The model preserving privacy performs better in the present system than the traditional models on one partitioning attribute.(Max depth=1).

Table 2: Comparison of the present homomorphic model to the traditional models on two partitioning attribute.(Max depth=2).

Instances	RSA	ECC	Homomorphic Encryption	Adaptive Homomorphic
#50	0.917	0.935	0.948	0.983
#100	0.913	0.922	0.944	0.977
#150	0.924	0.941	0.95	0.974
#200	0.919	0.933	0.944	0.979
#250	0.922	0.923	0.936	0.979
#300	0.914	0.922	0.936	0.983
#350	0.919	0.941	0.93	0.977
#400	0.921	0.94	0.927	0.976
#450	0.922	0.942	0.954	0.981
#500	0.914	0.944	0.954	0.984
#550	0.924	0.922	0.935	0.984
#600	0.924	0.928	0.937	0.979
#650	0.919	0.938	0.937	0.979
#700	0.921	0.934	0.945	0.983
#750	0.917	0.937	0.923	0.976
#800	0.918	0.945	0.925	0.981
#850	0.915	0.945	0.929	0.983
#900	0.916	0.922	0.93	0.975
#950	0.923	0.925	0.948	0.975
#1000	0.921	0.945	0.922	0.977

Table 2: explains the efficiency of the proposed model to conventional pattern preservation models for privacy. The model maintaining privacy in the present framework performs better on two partitioning attributes than the standard models.(Max depth=2).

Table 3: Comparison of the present homomorphic model to the traditional models on three partitioning attribute.(Max depth=3).

Instances	RSA	ECC	Homomorphic Encryption	Adaptive Homomorphic
#50	0.913	0.923	0.924	0.979
#100	0.922	0.941	0.951	0.984
#150	0.917	0.939	0.953	0.983
#200	0.915	0.943	0.936	0.977
#250	0.914	0.941	0.947	0.973
#300	0.913	0.941	0.926	0.977
#350	0.922	0.943	0.948	0.981
#400	0.921	0.936	0.948	0.981
#450	0.918	0.923	0.929	0.973
#500	0.92	0.926	0.926	0.976
#550	0.92	0.923	0.946	0.975
#600	0.915	0.937	0.945	0.973
#650	0.914	0.934	0.95	0.98
#700	0.915	0.937	0.953	0.984
#750	0.919	0.938	0.925	0.975

#800	0.92	0.931	0.942	0.972
#850	0.921	0.931	0.951	0.981
#900	0.913	0.93	0.923	0.981
#950	0.913	0.944	0.951	0.981
#1000	0.924	0.923	0.953	0.977

Table 3, explains the efficiency of the proposed model to conventional pattern preservation models for privacy. The model protecting privacy in the present method performs better on three partitioning properties than the standard models.(Max depth=3).

Table 4: Runtime computation of average partitioning based privacy preserving model to the traditional methods.

Instances	**RSA**	**ECC**	**Homomorphic Encryption**	**Adaptive Homomorphic**
#50	295	259	170	150
#100	601	500	339	314
#150	890	786	522	459
#200	1188	1042	682	541
#250	1529	1194	841	697
#300	1795	1541	1020	840
#350	2118	1840	1212	1013
#400	2385	2073	1387	1140
#450	2725	2278	1558	1408
#500	2996	2488	1706	1438
#550	3371	3018	1903	1501
#600	3672	2925	2083	1688
#650	3954	3079	2243	1935
#700	4246	3565	2433	1925
#750	4602	3664	2540	2042
#800	4770	4353	2715	2317
#850	5174	3989	2938	2583
#900	5393	4409	3132	2550
#950	5664	5025	3197	2666
#1000	5863	5264	3415	3066

Table 4: Explains the computational time of the present partitioning-based privacy preserving model on specific partitions to the conventional models. It is clearly observed from the table that the present model has less computational runtime compared to the current models.

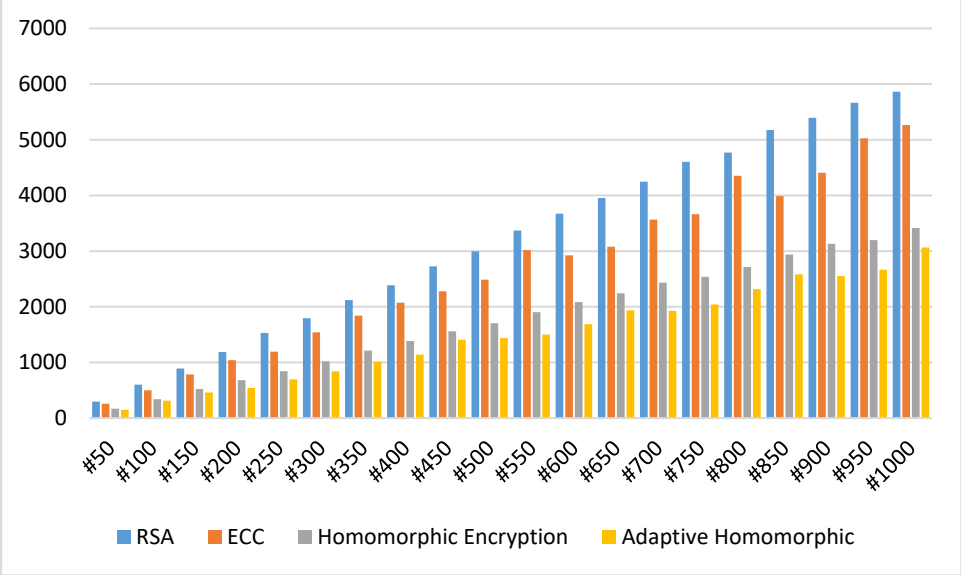


Figure 4: Describes the computational time of the present partitioning based privacy preserving model on different partitions to the traditional models. From the figure, the present model is clearly seen to have less computational runtime compared to the current models.

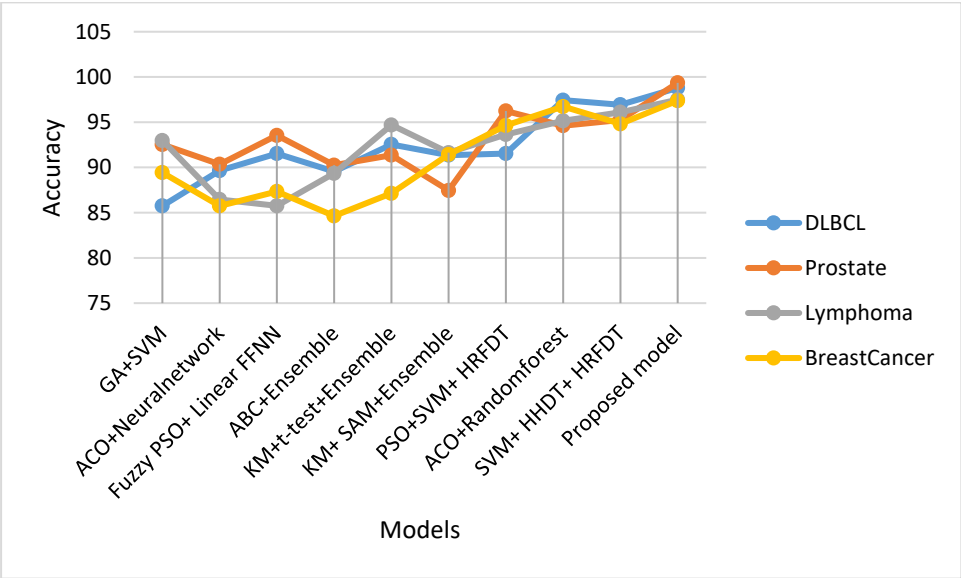


Figure 5: Performance analysis of proposed model to the existing models in terms of accuracy on different Microarray medical dataset.

Figure 5 illustrates the performance of the proposed deep learning framework to the traditional feature selection based classifiers in terms of data classification accuracy. From the figure, it is noted that the present deep learning framework is efficient than the existing models on large data variations.

CONCLUSION

Machine learning models such as grouping, clustering, or selection of features are applied to pattern analysis on the multiple datasets. For wide data sets the exposure of personal data or trends has become a major issue. Traditional privacy preserving deep learning frameworks are depends on data transformation approaches rather than the cryptographic approach due to high computational memory and time in cloud computing. In the real time multi-user applications, multiple datasets are distributed across the multiple users for privacy preserving. As the data size of multi-user applications increases, traditional PPDLM models require high computation memory and time for preserving the machine learning patterns. To overcome these problems, a novel data partitioning based privacy Preservation deep learning model is implemented on high dimensional datasets. Experimental results proved that the present system has high computational accuracy with privacy in the patterns compared to the existing models.

REFERENCES

1. Q. Wang, B. Qin, J. Hu, F. Xiao, Preserving transaction privacy in bitcoin, Fut. Gener. Comput. Syst. (2017), https://doi.org/10.1016/j.future.2017.08.026 in press.
2. F. Buccafurri, V.D. Angelis, G. Lax, S. Nicolazzo, A. Nocera, The challenge of privacy in the cloud, Encycl. Bioinform. Comput. Biol. 1 (2019) 265–271
3. M. Li, L. Lai, N. Suda, V. Chandra, D.Z. Pan, PrivyNet: A flexible framework for privacy preserving deep neural network training, (2018). arXiv:1709.06161v3[cs.LG].
4. S.A. Osia, A.S. Shamsabadi, A. Taheri, H.R. Rabiee, H. Haddadi, Private and scalable personal data analytics using hybrid edge-to-cloud deep learning, Computer 51 (5) (2018) 42–49.
5. H. Chabanne, A. Wargny, J. Milgram, Privacy-preserving classification on deep neural network, Proceedings of the Conference on Real World Cryptography, 2017, pp. 1–18.

6. P. Li, J. Li, Z. Huang, T. Li, C.Z. Gao, S.M. Yiu, K. Chen, Multi-key privacy-preserving deep learning in cloud computing, Fut. Gener. Comput. Syst. 74 (2017) 76–85.
7. S.A. Osia, A.S. Shamsabadi, A. Taheri, K. Katevas, H.R. Rabiee, N.D. Lane,H. Haddadi, Privacy-preserving deep inference for rich user data on the cloud, (2017). arXiv:1710.01727v3 [cs.CV].
8. W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, X. Chen, A sparse auto-encoder-based deep neural network approach for induction motor faults classification, Measurement 89 (2016) 171–178.
9. Marinelli, L., Castelletti, L., Trompetto, C.Isolated demyelination of corpus callosum following hypoxia(2018) European Journal of Molecular and Clinical Medicine, 5 (4), pp. 85-88. DOI: 10.5334/ejmcm.259
10. Anjum, A., Ahmad, N., Malik, S. U., Zubair & Shahzad, B. (2018). An efficient approach for publishing microdata for multiple sensitive attributes. The Journal of Supercomputing, 1-29.
11. Gong, Q., Luo, J., Yang, M., Ni, W., & Li, X. B. (2017). Anonymizing 1: M microdata with high utility. Knowledge-based systems, 115, 15-26.
12. J.W. Yuan, S.C. Yu, Privacy preserving back-propagation neural network learning made practical with cloud computing, IEEE Trans. Parallel Distrib Syst. 25 (1) (2015) 212–221