PSLP: Privacy-Preserving Single-Layer Perceptron Learning for e-Healthcare

Guoming Wang, Rongxing Lu, and Cheng Huang
School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798
Email: wang0947@e.ntu.edu.sg; rxlu@ntu.edu.sg; huangcheng@ntu.edu.sg

***Abstract*—In big data era, the explosive data mining techniques are used as popular tools to mine useful knowledge for the hospitals. However, considering the complexity of these techniques, the hospitals tend to outsource both data and calculations to computationally powerful cloud, which however poses a potential threat to user's privacy. In this paper, in order to address the privacy challenge, based on Paillier homomorphic cryptosystem, we propose a feasible privacy-preserving single-layer perceptron scheme, named PSLP. Specifically, in the proposed PSLP scheme, a hospital outsources the sensitive medical information to the cloud in ciphertext, and then the cloud can execute the privacy-preserving neural network training to obtain the disease model. Detailed security analysis shows the proposed PSLP can really achieve privacy-preserving property. In addition, extensive performance evaluations also demonstrate it is feasible in terms of computational cost and communication overhead.**

***Keywords:* Big data, e-Healthcare, single-layer perceptron neural network, privacy-preserving training**

I. Introduction

An era of open information in healthcare is now under way. Healthcare experts, service providers and other practitioners have stepped up to the plate and analyze big data to obtain insights [1]. Although these efforts are still in their early stages, they could collectively help the industry address problems relevant to the variability in healthcare quality and escalating healthcare spending. For instance, researchers can mine the data to see what treatments are most effective for particular conditions, gain important information that can help patients and reduce costs. Evaluation done by [2] revealed that over 200 businesses created since 2010 are developing a diverse set of innovative tools to make better use of available healthcare information. As the technological capabilities and understanding advance, innovators are expected to develop even more interesting ideas for using big data, some of which could help substantially reduce the soaring cost of healthcare [3].

Artificial neural networks (ANNs) could be beneficial as a tool for attention focusing and used to help building decision support models, which thus are able to alert both patients and healthcare practitioners of high risk of certain diseases or attributes. Tools like this can be a more cost-effective way of preliminary diagnosis that could or could not lead to additional and expensive diagnostic testing or variances in patient treatment plans [4]. For big-data initiatives to succeed, the healthcare system must undergo some fundamental changes. For example, stakeholders across the industry need to protect patient privacy as more information becomes public, and ensure that safeguards are in place to protect organizations from releasing information. In a nutshell, the privacy issue has become a big challenge for traditional ANNs in big data era.

In order to address the above privacy challenge and improve the accuracy of disease risk prediction in big data era, in this paper, we propose an efficient privacy-preserving single-layer perceptron learning scheme (PSLP), which is regarded as one of the widely used ANN algorithms. With the proposed PSLP, a hospital can outsource the encrypted medical data to a cloud server. Afterwards, the cloud server can calculate the disease model with these encrypted medical information without breaching privacy. Specifically, the main contributions of this paper are threefold.

- First, we propose PSLP, a privacy-preserving single-layer perceptron learning scheme, based on Paillier homomorphic cryptosystem [5]. With PSLP, a healthcare stakeholder, e.g., a hospital, can achieve privacy-preserving disease model training with the help of a computationally powerful cloud.
- Second, with detailed security analyses, we show that our proposed PSLP scheme can achieve the privacy-preservation for medical users. All the medical information stored on the cloud will not be leaked to the curious-but-honest cloud.
- Third, to validate the effectiveness of our proposed PSLP scheme, we also develop a cloud application and a hospital application in Java for our experiments, and evaluation results via extensive experiments show that our proposed PSLP scheme is privacy-preserving and feasible in big data era.

The remainder of this paper is organized as follows. In Section II, we introduce our system model, security model and design goal. In Section III, we recall some preliminaries for our scheme. Then, in Section IV, we present our PSLP scheme, followed by its security analyses and performance evaluation in Section V and Section VI respectively. We also discuss the related works in Section VII. Finally, we draw our conclusions in Section VIII.

II. Models and Design Goal

In this section, we formalize the system model, security model, and our design goal.

A. System Model

In our system model, we focus on the disease risk model training in the cloud with the help of a hospital. In such a way, our system model mainly includes two entities: a hospital , a cloud server, as shown in Fig. 1.

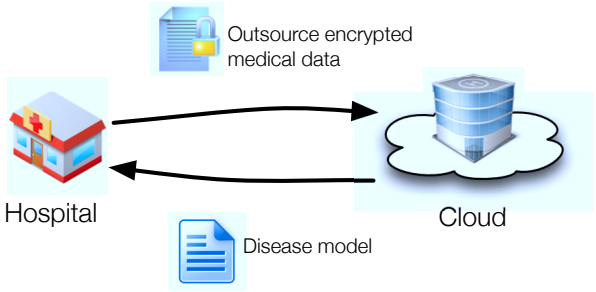

Fig. 1. System model under consideration

- Hospital: The hospital generates the private key and the public key at first and sends the public key to the cloud service provider. Then the hospital encrypts all the medical users' information and sends them to the cloud service provider. In the training procedure, the hospital will decrypt the processed medical information sent back from cloud, and feedback the sign $\{-1, 1\}$ of the result to the cloud. The cloud will terminate the training procedure if all the medical cases are matched or the iteration count is over a preset threshold.
- Cloud server: The cloud server is a curious but honest service provider with high computational capability and storage capability. The cloud trains the disease model with privacy-preserving single-layer perceptron learning protocol based on the encrypted medical information in a batch way. In each training iteration, the cloud sends the trained result to the hospital, and then the hospital decides to continue the training process or terminate it.

B. Security Model

In our security model, we consider the cloud service provider is *honest-but-curious*. That is, the cloud will faithfully follow the privacy-preserving single-layer perceptron learning protocol, but also attempt to know the sensitive medical information once the condition is satisfied. In specific, the disease prediction result of the training procedure is not privacy in our model, but the raw sensitive medical information needs privacy preservation.

Note that there may exist possible other attacks, i.e., forgery attack, in disease model training system. Since our focus is on privacy-preserving disease model training, those attacks are currently beyond the scope of this work, and will be discussed in future work.

C. Design Goal

Our design goal is to develop a secure and privacy-preserving single-layer perceptron learning scheme to provide disease risk model training service. Specifically, the following design goals should be guaranteed: *i) With our proposed scheme, the hospital can obtain a disease risk model from the cloud service provider without leaking the sensitive medical information. ii) The time-consuming computation will be executed in cloud, and all the data are stored in the cloud.*

III. Preliminaries

In this section, we review the Paillier Cryptosystem and outline the single-layer perceptron learning scheme, which will serve as the basis of our privacy-preserving single-layer perceptron learning scheme.

A. Paillier Cryptosystem

The Paillier Cryptosystem can achieve the homomorphic properties, which is widely desirable in many privacy-preserving applications. Concretely, the Paillier Cryptosystem is comprised of three algorithms: key generation, encryption and decryption.

• *Key Generation.* Given a security parameter k_1, two large prime numbers p_1, q_1 are first chosen, where $|p_1| = |q_1| = k_1$. Then, the RSA modulus $n = p_1\dot{q}_1$ and $\lambda = lcm(p_1 - 1, q_1 - 1)$ are computed. Define a function $L(u) = \frac{u-1}{n}$, after choosing a generator $g \in \mathbb{Z}^*_{n^2}$, $\eta = (L(g^\lambda \mod n^2))^{-1} \mod n$ is further calculated. Then, the public key is $pk = (n, g)$, and the corresponding private key is $sk = (\lambda, \eta)$.

• *Encryption.* Given a message $m \in \mathbb{Z}_n$, choose a random number $r \in \mathbb{Z}^*_n$, and the ciphertext can be calculated as $c = E(m) = g^m \cdot r^n \mod n^2$.

• *Decryption.* Given the ciphertext $c \in \mathbb{Z}_{n^2}$, the corresponding message can be recovered as $m = D(c) = L(c^\lambda \mod n^2)\dot{\eta} \mod n$. Note that, the Paillier Cryptosystem is provably secure against chosen plaintext attack, and the correctness and security can be referred to [5].

B. Single-layer Perceptron learning scheme

The single-layer perceptron (SLP), as introduced by Rosenblatt [6], is one of the earlier and simplest neural network architecture. The output layer is made of a single linear threshold unit with weights w that receives input x from a set of input neurons. That output unit uses the sign function as an activation function, thus implementing a two-class classification task onto the space {-1, 1}, where the class decision is based on the sign of the dot product $w^t x$.

• *Perceptron criterion.* We are given the training data $X = \{x_1, \ldots, x_M\}$ with associated desired outputs $\{t_1, \ldots, t_M\}$ (each $t_i \in \{-1, 1\}$) where -1 represents class C_1 and 1 represents C_2. For each input data x_i, the perceptron generates an output $o_i \in \{-1, 1\}$. Training minimizes the perceptron criterion $\mathbb{E}_p$ defined over the training set as $\mathbb{E}_p = -\sum_{i, o_i \neq t_i} t_i w^t x_i$, where the sum is performed over misclassified examples ($O_i \neq t_i$). This criterion attempts to find the linear hyperplane in the x-space that separates the two classes at the origin. To see this, observe that the dot product $-t_i w^t x_i$ is always positive for a misclassified x_i. Therefore, minimizing the perceptron criterion is equivalent to solving the linear inequality $t_i w^t x_i > 0$ for all x_i, which can be done either incrementally, one sample at a time, or in batch fashion.

IV. Proposed PSLP Scheme

In this section, we present our proposed PSLP scheme, which mainly consists of two phases: system setting and

privacy-preserving single-layer perceptron scheme, together with its correctness analysis.

A. System Setting

For a client-cloud single-layer perceptron learning system under consideration, the hospital bootstraps the whole system. Specifically, in the system initialization phase, given the security parameter k_1, the hospital calculates the Paillier Cryptosystem's public key $(n = p_1q_1, g)$, and the corresponding private key (λ, η), where p_1, q_1 are two large primes with $|p_1| = |q_1| = k_1$.

The medical information $x = \{x_1, \ldots, x_M\}$ are encrypted by the hospital and stored in the cloud service provider, while the associated desired outputs $\{t_1, \cdots, t_M\}$ (each $t_m \in \{-1, 1\}$) are stored in the cloud service provider in plain text. Because leaking $\{t_1, \cdots, t_M\}(t_m \in \{-1, 1\})$ is not harmful to the medical users, the security of this scheme will not be compromised. The encrypted medical information are stored as table I.

TABLE I
MEDICAL INFORMATION AND ASSOCIATED OUTPUT

Medical case	medical information	associated desired output
x_1	$\{x_{1,1}, x_{1,2}, \cdots, x_{1,D}\}$	t_1
x_2	$\{x_{2,1}, x_{2,2}, \cdots, x_{2,D}\}$	t_2
...		
x_M	$\{x_{M,1}, x_{M,2}, \cdots, x_{M,D}\}$	t_M

B. Privacy-preserving Single-layer Perceptron Learning

In the system setting, for each medical case x_i, the hospital has already encrypted it and stored it in the cloud. Therefore, the general training procedure of the single-layer perceptron learning can be described as follows:

- Step 1: The hospital randomly initializes a set of weight $w = (w_1, \cdots, w_d)$, in which not all w_i are equal to 0 and sends this weight vector $w = (w_1, \cdots, w_d)$ to the cloud.
- Step 2: The cloud will choose a piece of medical case $x_i = (x_{i,1}, \cdots, x_{i,d})$ by order.
- Step 3: With the weight vector $w = (w_1, \cdots, w_d)$ and the medical case $x_i = (x_{i,1}, \cdots, x_{i,d})$, the cloud computes the neuron output $o_i = sign(w^t \cdots x_i)$. If $o_i \neq t_i$, then cloud will update the weight vector $w = (w_1, \cdot, w_d)$ by the operation $w = w + \eta x_i t_i$, where η is the learning rate.
- Step 4: If all the training cases $x = \{x_1, \ldots, x_M\}$ are correctly satisfied or a specified minimum number of iterations are executed, the cloud will terminate the training algorithm and output the optimal set of weights $w = (w_1, \cdots, w_d)$. Otherwise, the cloud will choose the next medical case $x_{i+1} = (x_{i+1,1}, \cdots, x_{i+1,d})$, and repeat the steps from step 2.

However, the above general single-layer perceptron learning procedure does not address privacy issues. Therefore, in the following, we introduce a privacy-preserving single-layer perceptron learning scheme (PSLP) to ensure the cloud cannot get to know the medical information $x = \{x_1, \ldots, x_M\}$.

The main steps of PSLP, as shown in Fig. 2, are summarized as follows. Noticing that the medical information $x_{i,j} \in \vec{x_i} = (x_{i,1}, \cdots, x_{i,d})$ may be decimal originally. Here for the efficient computation in PSLP, each piece of medical information $x_{i,j}$ is expanded with $1,000$ times such that all $x_{i,j} \in \vec{x_i} = (x_{i,1}, \cdots, x_{i,d})$ are integer values lying in Z_n and $-Z_n$. All the weights $w_j \in \vec{w} = (w_1, \cdots, w_d)$ are also lying in Z_n and $-Z_n$. $ex_i = E(x_i)$ means the encrypted medical information $(E(x_{i,1}), \cdots, E(x_{i,d}))$ for the medical case $(x_{i,1}, \cdots, x_{i,d})$ and $ew_j = E(w_j)$ means the encrypted weight. The encryption algorithms for $x_{i,j}$ and w_j are as follows:

$$ex_{i,j} = \begin{cases} E(x_{i,j}), & x_{i,j} \geq 0 \\ E(n - |x_{i,j}|) & x_{i,j} < 0 \end{cases} \quad (1)$$

$$ew_j = \begin{cases} E(w_j), & w_j \geq 0 \\ E(n - |w_j|) & w_j < 0 \end{cases} \quad (2)$$

where n is the modulus number of the Paillier Cryptosystem's public key (n, g). The function sign(r) is defined as follows:

$$sign(r) = \begin{cases} 1, & \texttt{bitLength}(r) << \texttt{bitLength}(n) \\ -1, & \texttt{bitLength}(r) \approx \texttt{bitLength}(n) \end{cases} \quad (3)$$

Hospital: $\vec{w} = (w_1, \cdots, w_d)$, Cloud: $\{ex_1, \cdots, ex_M\}, \{t_1, \cdots, t_M\}, (x_m \in R^D, t_m \in \{-1, 1\})$, η

1.Hospital: sends $\vec{w}$ to cloud

2.Cloud: choose a medical case $e\vec{x}_i$
for $j = 1, \cdots, d$, compute $D_j = ex_{i,j}^{w_j}$, if $w_j >= 0$
or $D_j = ex_{i,j}^{n-|w_j|}$, if $w_j < 0$
end for
and compute $R = \prod_{j=1}^{d} D_j$
send the R to the hospital

3.Hospital: use Paillier decryption $Dec()$ to decrypt the R with private key, get the sign of the result,
$s = sign(Dec(R))$

4.Cloud: **if** $s \neq t_i$ and $t_i == 1, exp = \eta$
else if $s \neq t_i$and$t_i == -1, exp = n - \eta$
for $j = 1, \cdots, d$,
compute $u_j = ex_{i,j}^{exp}$,
compute $ew_j = E(w_j)$, compute $ew_j = ew_j \cdot u_j$
end for

5.Hospital: decrypt the vector $e\vec{w}$
if all the medical case are matched or training count
is larger than a threshold, terminate the algorithm
else: send the vector $\vec{w}$ to cloud and repeat step 2.

Fig. 2. The description of PSLP protocol

Step 1: The hospital randomly initializes a set of weight $\vec{w} = (w_1, \cdots, w_d)$ and sends the weight vector $\vec{w}$ to the cloud.

Step 2: After receiving the weight vector $\vec{w} = (w_1, \cdots, w_d)$, cloud chooses a piece of encrypted medical case $e\vec{x}_i = (ex_{i,1}, \cdots, ex_{i,d})$. For each $ex_{i,j} \in e\vec{x}_i$, the cloud computes $D_j = ex_{i,j}^{w_j}$, and then, the cloud computes the number $R = \prod_{j=1}^{d} D_j$. The result R is returned to the hospital.

Step 3: The hospital decrypts the R with the private key and gets the sign of the plain text $s = sign(Dec(R))$. Then the sign s is sent back to the cloud.

Step 4: Receiving the sign s, the cloud compares it with the associated output t_i. If s is not same as the associated output t_i and $t_i = 1$, with the learning rate η, cloud assigns the exponent

value exp with η. Or else, If s is not same as the associated output t_i and $t_i = -1$, the cloud assigns exp with $n-\eta$. With the exponent number exp, for each $ex_{i,j} \in e\vec{x}_i$, the cloud computes $u_j = ex_{i,j}^{exp}$. At the same time, all the weight in the vector $\vec{w}$ are encrypted as $ew_j = E(w_j)$. With the vector $\vec{u}$,the cloud updates the encrypted weight vector $e\vec{w}$ by operation for each weight $ew_j = ew_j \cdot u_j$, $j \in \{1, \cdots, d\}$. Finally, the cloud returns the updated vector $e\vec{w}$ to the hospital. Otherwise, if s is the same as the associated output t_i, the cloud will send back the vector $\vec{w}$ in plain text.

Step 5: If the received weight vector is the encrypted $e\vec{w}$, the hospital decrypts the vector $e\vec{w}$ and gets the weight vector $\vec{w}$. Then the hospital sends the weight vector $\vec{w}$ to the cloud, and repeats the step 2. The hospital will terminate the training procedure if all the medical cases are match, i.e., the received weight vector for all the medical cases $\{x_1, \cdots, x_M\}$ are in plain text. Or if the iteration count is larger than a threshold, the training is also terminated.

Correctness. The correctness of PSLP can be illustrated as follows: In the step 2, when $x_{i,j} >= 0, w_j >= 0$, the cloud computes:

$$\begin{aligned} D_j = ex_{i,j}^{w_j} = E(x_{i,j})^{w_j}, \text{ for } j = 1, \cdots, d \\ = E(x_{i,j} \cdot w_j) \end{aligned} \tag{4}$$

Then the cloud computes R, where

$$R = \prod_{j=1}^{d} D_j = \prod_{j=1}^{d} E(x_{i,j} \cdot w_j) = E\left(\sum_{j=1}^{d} x_{i,j} \cdot w_j\right) \tag{5}$$

In the step 3, Hospital receives R and decrypts it to get s, where

$$\begin{aligned} s = sign(Dec(R)) = sign(Dec(E(\sum_{j=1}^{d} x_{i,j} \cdot w_j))) \\ = sign(\sum_{j=1}^{d} x_{i,j} \cdot w_j) \end{aligned} \tag{6}$$

When $\sum_{j=1}^{d} x_{i,j} \cdot w_j >= 0$, the bit length of $\sum_{j=1}^{d} x_{i,j} \cdot w_j$ is far less than that of n, thus s is 1. On the other hand, when $\sum_{j=1}^{d} x_{i,j} \cdot w_j < 0$, the bit length of $\sum_{j=1}^{d} x_{i,j} \cdot w_j$ is almost the same as that of n, therefore, s is -1.

Giving the condition $s \neq t_i$, in the step 4, when $x_{i,j} >= 0, \eta >= 0$, the cloud computes:

$$u_j = ex_{i,j}^{exp} = ex_{i,j}^{\eta} = E(x_{i,j} \cdot \eta), \text{ for } j = 1, \cdots, d \tag{7}$$

After encrypting the old weight vector, the new weight vector is updated as follows:

$$\begin{aligned} ew_j = ew_j \cdot u_j = E(w_j) \cdot E(x_{i,j} \cdot \eta), \text{ for } j = 1, \cdots, d \\ = E(w_j + x_{i,j} \cdot \eta) \end{aligned} \tag{8}$$

Same as the above, either $x_{i,j} < 0$ or $w_j < 0$, the cloud can correctly compute R and hospital can correctly calculate s. Noticing that, when $x_{i,j}$ is less than 0, $x_{i,j}$ will be encrypted as:

$$ex_{i,j} = E(n - |x_{i,j}|), \text{ for } j = 1, \cdots, d \tag{9}$$

The same as $x_{i,j}$, when w_j is less than 0, w_j is encrypted as:

$$ew_j = E(n - |w_j|), \text{ for } j = 1, \cdots, d \tag{10}$$

From the above observations, the hospital can get disease model $\vec{w}$ with the help of the cloud. As a result, the correctness of the PSLP protocol is satisfied.

V. Security Analysis

In this section, we analyze the security of our proposed PSLP scheme. Especially, we focus on how the proposed PSLP can achieve the privacy-preservation of medical users' information $x = \{x_1, \ldots, x_M\}$.

• *Security of the medical information stored in the cloud.* In the proposed scheme, medical data $x = \{x_1, \ldots, x_M\}$ offered by the hospital are all encrypted by the Paillier Cryptosystem generated by the hospital. For each medical case $x_m = (x_1, \cdots, x_d)$, the encrypted form can be expressed as $E(x_m) = (g^{x_{m,1}} r^n \mod n^2, \cdots, g^{x_{m,d}} r^n \mod n^2)$. Since Paillier Cryptosystem is semantic secure against the chosen plaintext attack, the medical data are semantic secure and privacy-preserving. Therefore, although all the data are stored in the cloud, the cloud can not identify the corresponding contents.

• *Security of the medical information in the PSLP scheme.* In the step 2, the cloud do computation $D_j = ex_(i,j)^{w_j} = (g^{x_{i,1} w_j} r^{n w_j} \mod n^2$ as $w_j >= 0$, or $D_j = ex_(i,j)^{n-w_j} = (g^{x_{i,1}(n-w_j)} r^{n(n-w_j)} \mod n^2$ as $w_j < 0$. All these computations are operated over ciphertext, thus, all the medical information are privacy-preserving. Similiarly, the computations in the step 4 are privacy-preserving for all the medical data.

From the above analyses, we can see that the proposed PSLP scheme is secure and privacy-preserving, and can achieve our security design goal.

VI. Performance Evaluation

In this section, we evaluate our proposed PSLP scheme in terms of computational cost and communication overhead.

In specific, in our experiments, we develop our PSLP in Java. The database we used can be found on [7]. The name of the dataset is Wisconsin Breast Cancer Database (January 8, 1991), which was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg. The attributes information include: clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli, mitoses. The domain of all these attributes are integers from 1 to 10. We set the class attribute benign as 1, while malignant as -1. Because some of these attributes are empty (question mark in the dataset), we choose the medical cases with full attributes. In addition, we constructed an instance of the Paillier cryptosystem with 1024 bits of modulus and at least $1 - 2^{-64}$ certainty of primes generation. We set the iteration threshold as 10000, and the experimental results show that the error rate for the whole dataset is around 17.4%, which is acceptable.

Computational cost. Table II illustrates the computational costs for each step of the PSLP scheme. For expression simplicity, in the table, we denote time cost of one exponentiation operation on ciphertext in Paillier cryptosystem as EXP, that of one multiplication operation on ciphertext as MUL and that of one modular inverse operation in the decryption algorithm in Paillier cryptosystem as DIV. In the step 2, d exponents are executed to compute the $D_j, j \in (1, \cdots, d)$, and $d-1$ EXP for the value R. In the step 3, for the hospital, only one decryption is executed, which includes one EXP and one DIV. In addition, d EXP are executed to compute the $u_j, j \in (1, \cdots, d)$, d extra EXP and d MUL are executed for the encryption of weight vector $\vec{w}$ and d MUL to update the vector $\vec{w}$. At the last step, the hospital executes d EXP and d DIV to decrypt the updated weight vector $\vec{w}$. It is obvious that most of the time-consuming operations are executed in the cloud side, so that it is practical.

TABLE II
FOR EACH MEDICAL CASE x_i, THE COMPUTATIONAL COST

Step	Hospital, Cloud	Computation Cost
step 1	–	–
step 2	Cloud	d EXP + $d-1$ MUL
step 3	Hospital	1 EXP + 1 DIV
step 4	Cloud	$2d$ EXP + $2d$ MUL
step 5	Hospital	d EXP + d DIV

Communication overhead. Table III demonstrates the communication overhead for our proposed PSLP scheme. Based on the above parameter setting, the length of weight vector $\vec{w}$ is at most $1024d$ bits, the length of R is at most 1024 bits, and the length of the encrypted updated weight vector $e\vec{w}$ is $2048d$ bits. Note that, the communication overhead can be reduced with the batch PSLP algorithm, which will be discussed in future work.

TABLE III
THE COMMUNICATION OVERHEAD

Message	Communication Overhead (bit)
$\vec{w}$	$1024 \cdot d$
R	1024
$e\vec{w}$	$2048 \cdot d$

VII. RELATED WORKS

In this section, we briefly discuss some related works on disease model and privacy-preserving neural network learning system, which are closely related to our work. As early diagnosis of disease can minimize the side-effects, safety risks and the financial costs, *disease risk prediction* has attracted considerable attention. For example, in 2012, Anooj et al. [8] develop a fuzzy rule-based decision support system for prediction of heart disease. In 2013, Bouwmeester et al. [9] use the multivariate logistic regression technique for developing the risk prediction model, in which a linear combination of predictors associated with multiple symptoms and environmental data are used to fit a logarithmic transformation of the probability of the tested disease. Actually, to flourish the medical industry, more and more data analysis companies are encouraged to mine new knowledge for efficient medical service.

Neural network systems are highly capable of deriving knowledge from complex data, and they are used to extract patterns and trends which are otherwise hidden in many applications [10], [11]. Saeed Samet and Ali Miri [10] proposeda privacy-preserving protocols for back-propagation and extreme learning machine algorithms when data is horizontally and vertically partitioned among several parties. For more big-data initiatives to succeed, the healthcare system must undergo some fundamental changes. For example, stakeholders across the industry need to protect patient privacy. Our proposed neural network training algorithm is based on Paillier encryption system [5]. All the data are computed on ciphertext, which perfectly achieves the privacy-preserving requirement in healthcare system .

VIII. CONCLUSIONS

In this paper, we have proposed an efficient and privacy-preserving single-layer perceptron scheme, called PSLP. The proposed PSLP is characterized by employing a hormomorphic paillier cryptosystem, which enables the outsourced medical data are processed on cloud in ciphertext without leaking the sensitive medical information. Detailed analysis shows the PSLP really achieves the training target and gets the disease model. In future work, we aim at dealing with more efficient and privacy-preserving big-data medical model training algorithm.

REFERENCES

[1] R. Lu, H. Zhu, X. Liu, J. K. Liu, and J. Shao, "Toward efficient and privacy-preserving computing in big data era," *IEEE Network*, vol. 28, no. 4, pp. 46–50, 2014.

[2] P. Groves, B. Kayyali, D. Knott, and S. Van Kuiken, "The big datarevolution in healthcare," *McKinsey Quarterly*, 2013.

[3] G. Wang, R. Lu, and C. Huang, "Pguide: An efficient and privacy-preserving smartphone-based pre-clinical guidance scheme," in *IEEE Globecom'15*, San Diego, CA, USA, December 6 - 10 2015.

[4] J. Nolting. Developing a neural network model for health care. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1839654/

[5] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Advances in Cryptology - EUROCRYPT '99*, 1999, pp. 223–238.

[6] N. J. Bershad, N. Cubaud, and J. J. Shynk, "Stochastic convergence analysis of the single-layer backpropagation algorithm for noisy input data," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1315–1319, 1996. [Online]. Available: http://dx.doi.org/10.1109/78.502354

[7] M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml

[8] P. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules," *Journal of King Saud University-Computer and Information Sciences*, vol. 24, no. 1, pp. 27–40, 2012.

[9] W. Bouwmeester, J. W. Twisk, T. H. Kappen, W. A. Klei, K. G. Moons, and Y. Vergouwe, "Prediction models for clustered data: comparison of a random intercept and standard regression model," *BMC medical research methodology*, vol. 13, no. 1, p. 19, 2013.

[10] S. Samet and A. Miri, "Privacy-preserving back-propagation and extreme learning machine algorithms," *Data & Knowledge Engineering*, vol. 79, pp. 40–61, 2012.

[11] J. Yuan and S. Yu, "Privacy preserving back-propagation neural network learning made practical with cloud computing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 1, pp. 212–221, 2014. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/TPDS.2013.18