An Efficient and Privacy-Preserving Physiological Case Classification Scheme for E-healthcare System

Gang Shen[1], Yumin Gui[2(✉)], Mingwu Zhang[1,3,4], Yu Chen[1], Hanjun Gao[5], and Yixin Su[6]

[1] School of Computer Science, Hubei University of Technology, Wuhan 430068, China
shengang@hbut.edu.cn, csmwzhang@gmail.com, ychen@hbut.edu.cn
[2] Department of Ophthalmology, Wuhan Puren Hospital, Wuhan 430080, China
5537468@qq.com
[3] School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541004, China
[4] Hubei Key Laboratory of Intelligent Geo-Information Processing, China University of Geosciences, Wuhan 430074, China
[5] China Nuclear Power Operation Technology Corporation, LTD, Wuhan 430070, China
gaohj@cnnp.com.cn
[6] School of Automation, Wuhan University of Technology, Wuhan 430070, China
suyixin@whut.edu.cn

Abstract. In this work, an efficient and privacy-preserving physiological case classification scheme for e-healthcare system (EPPC) is proposed. Specifically, a homomorphic cryptosystem combined with a support vector machine (SVM) algorithm is applied to efficiently classify the physiological cases without compromising patients' privacy. In terms of the EPPC, it has the capability of diagnosing the patient's symptom in a timely manner. In addition, a signature authentication technology applied in EPPC can efficiently prevent data from being forged or modified. Security analysis result shows that the proposed EPPC scheme has the following advantages: protect the privacy of patients; ensure that the classification parameters of SVM are secured. Compared with the existing works, the proposed EPPC scheme shows significant advantages in terms of computational costs and communication overheads.

Keywords: E-healthcare system · Privacy protection · Physiological case classification · Homomorphic cryptosystem

1 Introduction

With the rapid increase of the aging population, the limited medical resources are far from satisfying the high requirement of medical services in quality [3,6,13].

D. Wang et al. (Eds.): SPNCE 2020, LNICST 344, pp. 172–185, 2021.
https://doi.org/10.1007/978-3-030-66922-5_12

Fortunately, e-healthcare, a new solution and one of the latest popular research fields, can effectively meet the demand for health monitoring and the limited medical resources. As a part of e-healthcare system, wireless body area network (WBAN), a popular technology, has the capability of diagnosing and monitoring of the patient's physical health in a real-time manner.

The e-healthcare system has following benefits [4,14]: i) provide remote health monitoring and real-time diagnosis for patients; ii) realize the online communication between patients and physicians; iii) improve the efficiency of medical treatment, and reduce the cost of medical treatment. However, security and privacy problems in e-healthcare system also pose enormous challenges [10,19]. For example, with these physiological data, a rogue is easier to deduce the physical condition of the patients, which may cause further psychological and physical harm to the patients. Therefore, protecting patients' privacy is imperative in the e-healthcare system.

Support vector machine (SVM), a machine learning tool, is commonly used in various fields, such as disease prediction [16], face recognition [8], text classification [11], handwriting recognition [12] and bioinformatics [2] etc. In general, training and testing are two parts of SVM [9]. The classification parameters of SVM classifier are obtained by creating the characteristics of different types of data sets in the training phase [5]. In the testing phase, all unlabeled data samples are classified by the classification parameters of SVM classifier and marked as a matching class. Therefore, classifiers with clinical data sets are used to identify patient data in medical diagnosis system in the testing phase. In this paper, the SVM is used to solve the problem of patient physiological case classification in the e-healthcare system.

In view of the sensitivity of information (e.g., patients' health information, medical institution information, etc.) in the process of e-health medical diagnosis, it is imperative to save the patients' physiological data and the classification parameters of the SVM classifier in healthcare centre (HC). In other words, the sensitive data in e-healthcare system shall be fully protected and shall not be compromised. Although numerous up-to-date clinical decisions based on classification schemes are proposed [15–18], the SVM tools are rarely used in the machine learning. Even if the SVM tools are used, these schemes have the disadvantages of large computation cost. What's worse, the method of obtaining the sign of an encrypted value is complicated, and the computational cost is also high. Therefore, the challenge in designing a secure and efficient scheme for physiological case classification is significant in the research field.

In this work, we propose an EPPC scheme for e-healthcare system, enabling to classify the patients' data without compromising their privacy, thereby protecting the security of the e-healthcare system. Specifically, the **main contributions** of this work are as fourfold.

(1) Firstly, we propose an EPPC scheme for e-healthcare system by combining Okamoto-Uchiyama (OU) homomorphic cryptosystem technology and SVM in machine learning. In the proposed EPPC scheme, the homomorphic properties of the OU encryption scheme can be directly used to implement

operations on the ciphertext data, and SVM algorithm is applied to obtain the accurate diagnostic results. Therefore, it protects the privacy of patient user (PU) and HC and also helps PU obtain the results of online diagnosis quickly.

(2) Secondly, with the method of scaling variables in scheme [9] for reference, we successfully solve the following problems: OU scheme only supports integers; system variables can only be continuous. In addition, the proposed EPPC scheme also applies BLS [1] signature technology to prevent the PU's physiological data from being forged or modified.

(3) Thirdly, we develop a novel method to obtain a sign of the ciphertext of classification label. Because it is a lightweight method, it can significantly decrease the computation cost of HC in our scheme.

(4) Finally, in order to demonstrate the efficiency of EPPC, we build a simulator in JAVA and compare the performance of EPPC with related scheme. The comparison results show that our scheme is more efficient.

The rest of the paper is structured as follows. In Sect. 2, we present the system model and security requirements of this paper. Then, we introduce the preliminary knowledge in Sect. 3. The concrete EPPC scheme is proposed in Sect. 4. In Sect. 5, the security analysis is described. Next, we illustrate the performance evaluation in Sect. 6. Finally, we summarize this paper in Sect. 7.

2 System Model and Security Requirements

In this section, we will present the system model, security requirements and design goals related to this paper.

2.1 System Model

The entities in system model include a healthcare centre and a patient user, as shown in Fig.1. The description is detailed as follows.

Healthcare centre (HC): HC is an incompletely trusted entity which can normally run the protocol as specified one but may try to learn about the maximum physiological data from the protocol under the influence of the adversaries. There are a large number of case training sets in the database of HC. HC is responsible for helping PU classify the physiological data, thus diagnosing his/her disease condition.

Patient user (PU): PU's physiological data and personal information can be sent to HC through the sensor devices. The physiological data is encrypted by user before it is sent to protect the patient's privacy.

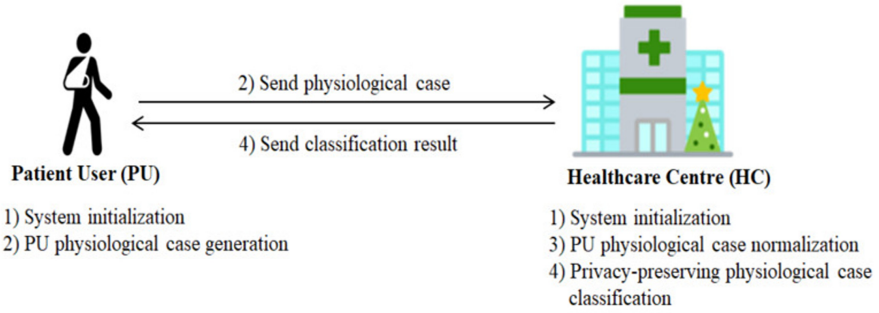


Fig. 1. System model of EPPC under consideration.

2.2 Security Requirements

The proposed EPPC scheme mainly meets the security requirements of the following three aspects.

(1) *Privacy.* The physiological data, reflecting PU's health status, involves PU's privacy, which should be protected. In other words, even if HC obtains PU's physiological case, it cannot identify PU's physiological data. Moreover, even if the adversary $\mathcal{A}$ obtains PU's physiological data from HC' databases, he/she still does not know about PU's health status.
(2) *Authentication.* The authentication of PU's physiological case enables HC to know whether the physiological case is sent by a valid PU, thereby preventing the physiological case from being modified or forged by adversary $\mathcal{A}$ during the transmission.
(3) *Confidentiality.* Since the training set data and the parameters of the SVM classification in HC are obtained by spending a lot of time and financial resources, these parameters are confidential data for HC. Even if PU receives the diagnosis result from HC, he/she will not learn any information of these parameters.

3 Preliminaries

3.1 Bilinear Pairing

Let $\mathbb{G}_1$, $\mathbb{G}_2$ be two cyclic groups of the same prime order q and $e : \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$ is a bilinear map that satisfies the properties as following:

(1) *Bilinearity.* $e(aP, bQ) = e(P, Q)^{ab}$ for all $a, b \in \mathbb{Z}_q^*$, and $P, Q \in \mathbb{G}_1$ are generators.
(2) *Non-degeneracy.* $e(P, Q) \neq 1$.
(3) *Computability.* $e(P, Q)$ can be computed by an efficient algorithm.

Definition 1. Let $\mathcal{G}en$ be a probabilistic algorithm, and input a security parameter κ, then output $(q, P, \mathbb{G}_1, \mathbb{G}_2, e)$, where q is a κ-bit prime number.

Definition 2. (Computational Diffie-Hellman (CDH) Problem). The CDH problem can be described as follows: For $a, b \in \mathbb{Z}_q^*$, given $(P, aP, bP) \in \mathbb{G}_1$, compute $abP \in \mathbb{G}_1$.

3.2 Support Vector Machine (SVM)

SVM, a two-class or multi-class classification model, is used to process data classification [9], which attracts wide attention and has been used for a long time since its birth. The liner classifier with the largest interval in the feature space can be used as the basic model of SVM. However, most of the data are not linearly divisible. In that way, it is necessary to map the sample from the original space to a higher dimensional feature space with the help of kernel function, so that the sample can be linearly separated in this feature space.

For linear classification problem, the training samples are linearly separable, so the decision function is

$$d(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = \sum_{i \in S} \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b, \tag{1}$$

where α_i are Lagrangian variables and $\mathbf{x}_i \in S$ are support vectors for $i = 1, \cdots, |S|$, S is the set of support vectors. In addition, $\mathbf{w}$ and b are the classification parameters, and $y_i \in \{+1, -1\}$ is the classification label of sample $\tilde{\mathbf{x}}_i$ for $i = 1$ to $|S|$. For non-linear classification problem, the training samples are not linearly inseparable, then the dot product (i.e., $\mathbf{x}_i^T\mathbf{x}$) in Eq. 1 should be substituted by different kernel functions. So the decision function can be modified as

$$d(\mathbf{x}) = \sum_{i \in S} \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}_j) + b. \tag{2}$$

In general, the samples with high feature dimension are linearly divisible, and the linear kernel function can be considered. Because the dimension of the sampled data in our scheme is high, we only consider a linear kernel in this work. Therefore, the decision function is

$$d(\mathbf{x}) = \sum_{i \in S} \alpha_i y_i (\mathbf{x}_i^T \mathbf{x}_j) + b. \tag{3}$$

3.3 Okamoto-Uchiyama (OU) Homomorphic Cryptosystem

The OU cryptosystem, proposed by Okamoto and Uchiyama [7], supports the additive homomorphism. In the OU cryptosystem, when the security parameter is 512 bits, the size of plaintext and ciphertext are approximately 512 bits and 1536 bits, respectively. The OU cryptosystem includes the following three algorithms:

(1) *Key generation.* Given the security parameter κ', select two large primes p' and q' with the same length $|p'| = |q'| = \kappa'$, and calculate $N = p'^2 q'$.

Then, choose $g \in \mathbb{Z}_N^*$ such that $g^{p'} \neq 1 \ mod \ p'^2$, and let $h = g^N mod \ N$. $pk = (N, g, h)$ and $sk = (p', q')$ are the public key and private key of the cryptosystem, respectively.

(2) *Message encryption.* Choose a random number $r \in \mathbb{Z}_N$, and calculate the ciphertext $C = E(m) = g^m \cdot h^r \ mod \ N$, where m is a message, $0 \leq m < 2^{\kappa'-1}$.

(3) *Message decryption.* Give the ciphertext $C \in \mathbb{Z}_N$, the corresponding message can be decrypted by calculating $m = D(C) = \frac{L(C^{p'-1} \ mod \ p'^2)}{L(g^{p'-1} \ mod \ p'^2)} \ mod \ p'$, where $L(x) = \frac{x-1}{p'}$.

In addition, the OU cryptosystem satisfies the additive homomorphism, and the specific form is as follow:

$$\begin{aligned} D(E(m_1) \cdot E(m_2) \ mod \ N) &= D(g^{m_1} h^{r_1} \cdot g^{m_2} h^{r_2} \ mod \ N) \\ &= D(g^{m_1+m_2} h^{r_1+r_2} \ mod \ N) \\ &= D(E(m_1 + m_2) \ mod \ N) \end{aligned}$$

where $0 \leq m_1 + m_2 < 2^{\kappa'-1}$.

4 The Proposed EPPC Scheme

In this section, we give the concrete EPPC scheme, which consists of five parts: system initialization, PU physiological case generation, PU physiological case normalization and privacy-preserving physiological case classification. In order to make reader have a better understanding, we only consider the classification of one PU's physiological case by HC in our work.

In EPPC scheme, we assume that HC owns the case training set of points $\tilde{\mathbf{t}}_i \in \mathbb{R}^n$, $i = 1, \cdots, n$, where $\tilde{\mathbf{t}}_i = (\tilde{t}_{i1}, \tilde{t}_{i2}, \cdots, \tilde{t}_{in})$ and each point $\tilde{\mathbf{t}}_i$ belongs to one of the two classes denoted by the label $y_i \in \{-1, +1\}$, $i = 1, \cdots, n$. Therefore, we can use these case samples to train SVM to classify the unlabeled test sample. In addition, the training case shall be normalized to keep their values on the same scale to prevent a large original samples from biasing the solution [10]. The normalized training case samples can be denoted as

$$\mathbf{t}_i = \frac{\tilde{\mathbf{t}}_i - \bar{\mathbf{t}}}{\sigma^2}, \tag{4}$$

where $\mathbf{t}_i \in \mathbb{R}^n$, $i = 1, \cdots, n$, σ and $\bar{\mathbf{t}}$ denote the standard deviation and the mean of the training case samples, respectively.

PU in our scheme means undiagnosed patient who has an n-dimensional physiological data $\tilde{\mathbf{d}} = (\tilde{d}_1, \tilde{d}_2, \cdots, \tilde{d}_n)$ received by sensors, where $\tilde{\mathbf{d}} \in \mathbb{R}^n$, $i = 1, \cdots, n$. These elements in the vector $\tilde{\mathbf{d}}$ can be expressed as different physiological indicators of PU. For example, d_1 represents blood sugar, d_2 denotes blood pressure, d_3 means blood lipids, etc. In this work, PU encrypts its physiological data and uploads them to HC's data server. In that case, the data normalization should be carried out in the form of ciphertext, and the specific process can be found in Sect. 4.3.

4.1 System Initialization

Given the security parameter κ, PU generates $(q, \mathbb{G}_1, \mathbb{G}_2, P, e)$ by running a key generation algorithm $\mathcal{G}en(\kappa)$, and then computes the OU cryptosystem's public key $pk = (N = p'^2q', g, h)$ and the corresponding private key $sk = (p', q')$, where p' and q' are two large primes with the same length $|p'| = |q'| = \kappa'$. PU also chooses a random number $x_u \in \mathbb{Z}_q^*$ as his/her signature private key, and computes the corresponding public key $Y = x_uP$. Additionally, PU chooses a cryptographic hash function H, where $H : \{0,1\}^* \rightarrow \mathbb{G}_1$.

Finally, PU exposes the common parameters $(q, \mathbb{G}_1, \mathbb{G}_2, P, e, N, g, h, Y, H)$, and keeps the private key (p', q', x_u) secretly.

HC randomly chooses a sufficiently large positive number ξ as a scaling factor to scale the decision function, which enables all the variables contained in the decision function to be quantized to the nearest integer value. HC also randomly selects two integers A and B to meet the following two conditions: i) $A > B$ and ii) $|A \cdot d(\mathbf{d}) + B| < 2^{\kappa'-2}$, where $|d(\mathbf{d})| < 2^l$ and $l \in \mathbb{Z}_N$. Since the data in our scheme is high dimensional, HC chooses a decision function with linear kernel function as

$$d(\mathbf{d}) = \sum_{i \in S} \alpha_i y_i \mathbf{t}_i^T \mathbf{d} + b, \tag{5}$$

where α_i are Lagrangian variables, $y_i \in \{+1, -1\}$ is the classification label of sample $\tilde{\mathbf{t}}_i$ for $i = 1$ to $|S|$, $\mathbf{t}_i$ and $\mathbf{d}$ are normalized training case samples for $i = 1, \cdots, |S|$ and normalized PU's physiological data, respectively. Here, $(\alpha_i, y_i, \mathbf{t}_i, b)$ are the classification parameters.

In the end, HC keeps $(\alpha_i, y_i, \mathbf{t}_i, b, \xi, A, B)$ secretly.

4.2 PU Physiological Case Generation

PU collects the physiological data $\tilde{\mathbf{d}} = (\tilde{d}_1, \tilde{d}_2, \cdots, \tilde{d}_n)$ through implantable, wearable or environmental sensor devices. The physiological data should be encrypted before they are sent to HC. Specifically, PU encrypts each element of the physiological data $\tilde{\mathbf{d}}$ individually using OU homomorphic cryptosystem to generate PU's physiological case $[\![\tilde{\mathbf{d}}]\!]$. The specific process is as follows:

(1) PU selects a random number $r_i \in \mathbb{Z}_N$, and calculates

$$\begin{aligned} [\![\tilde{d}_1]\!] &= g^{\tilde{d}_1} \cdot h^{r_1} mod N, \\ [\![\tilde{d}_2]\!] &= g^{\tilde{d}_2} \cdot h^{r_2} mod N, \\ &\cdots\cdots \\ [\![\tilde{d}_n]\!] &= g^{\tilde{d}_n} \cdot h^{r_n} mod N, \end{aligned}$$

where $i = 1, \cdots, n$. Then, the PU's physiological case is generated as

$$\begin{aligned} [\![\tilde{\mathbf{d}}]\!] &= (g^{\tilde{d}_1} \cdot h^{r_1} mod N, g^{\tilde{d}_2} \cdot h^{r_2} mod N, \cdots, g^{\tilde{d}_n} \cdot h^{r_n} mod N) \\ &= ([\![\tilde{d}_1]\!], [\![\tilde{d}_2]\!], \cdots, [\![\tilde{d}_n]\!]). \end{aligned} \tag{6}$$

(2) Use the private key x_u to generate a signature δ as

$$\delta = x_u \cdot H([\![\tilde{\mathbf{d}}]\!]||Y||T), \tag{7}$$

where T denotes PU's current timestamp.
(3) Send PU's physiological case $[\![\tilde{\mathbf{d}}]\!]||Y||T||\delta$ to HC.

4.3 PU Physiological Case Normalization

Because the OU cryptosystem does not support non-integers and the test samples and variables in SVM classification are continuous values [10], PU's physiological case should be quantized to the nearest integer value by HC before classification. The specific method is similar to Eq. 4.

(1) After receiving PU's physiological case $[\![\tilde{\mathbf{d}}]\!]||Y||T||\delta$, HC first verifies the validity of the timestamp and signature. If $e(P,\delta) = e(Y, H([\![\tilde{\mathbf{d}}]\!]||Y||T))$ does hold, the signature is accepted. Since $Y = x_u P$ and $\delta = x_u \cdot H([\![\mathbf{d}]\!]||Y||T)$, $e(P,\delta) = e(P, x_u \cdot H([\![\tilde{\mathbf{d}}]\!]||Y||T)) = e(Y, H([\![\tilde{\mathbf{d}}]\!]||Y||T))$.
(2) Using the parameters such as mean $\bar{\mathbf{t}}$ and standard deviation σ of training case samples and scaling factor ξ, HC can compute the values $\{(p'-1)\frac{\xi \bar{t}_1}{\sigma^2}, \cdots, (p'-1)\frac{\xi \bar{t}_n}{\sigma^2}\}$.
(3) Since $(-1)\ mod\ p' = (p'-1)\ mod\ p'$, HC encrypts each of $\{(p'-1)\frac{\xi \bar{t}_1}{\sigma^2}, \cdots, (p'-1)\frac{\xi \bar{t}_n}{\sigma^2}\}$ by OU homomorphic cryptosystem as $\{[\![(-1)\frac{\xi \bar{t}_1}{\sigma^2}]\!], \cdots, [\![(-1)\frac{\xi \bar{t}_n}{\sigma^2}]\!]\}$.
(4) After verifying the signature, depending on the homomorphic property, HC computes the normalized value of each element in $\{[\![\tilde{d}_1]\!], [\![\tilde{d}_2]\!], \cdots, [\![\tilde{d}_n]\!]\}$ individually with $\bar{\mathbf{t}}$, σ and ξ as follows:

$$\begin{aligned}
[\![\xi d_1]\!] &= [\![\tilde{d}_1]\!]^{\frac{\xi}{\sigma^2}} \cdot [\![(-1)\frac{\xi \bar{t}_1}{\sigma^2}]\!] = [\![\tilde{d}_1]\!]^{\frac{\xi}{\sigma^2}} \cdot [\![\bar{t}_1]\!]^{(-1)\frac{\xi}{\sigma^2}} \\
&= ([\![\tilde{d}_1]\!] \cdot [\![\bar{t}_1]\!]^{-1})^{\frac{\xi}{\sigma^2}} = [\![\frac{\tilde{d}_1}{\sigma^2} - \frac{\bar{t}_1}{\sigma^2}]\!]^{\xi} \\
&= [\![\frac{\tilde{d}_1 - \bar{t}_1}{\sigma^2}]\!]^{\xi} \\
[\![\xi d_2]\!] &= [\![\tilde{d}_2]\!]^{\frac{\xi}{\sigma^2}} \cdot [\![(-1)\frac{\xi \bar{t}_2}{\sigma^2}]\!] = [\![\tilde{d}_2]\!]^{\frac{\xi}{\sigma^2}} \cdot [\![\bar{t}_2]\!]^{(-1)\frac{\xi}{\sigma^2}} \\
&= ([\![\tilde{d}_2]\!] \cdot [\![\bar{t}_2]\!]^{-1})^{\frac{\xi}{\sigma^2}} = [\![\frac{\tilde{d}_2}{\sigma^2} - \frac{\bar{t}_2}{\sigma^2}]\!]^{\xi} \\
&= [\![\frac{\tilde{d}_2 - \bar{t}_2}{\sigma^2}]\!]^{\xi} \\
&\cdots\cdots \\
[\![\xi d_n]\!] &= [\![\tilde{d}_n]\!]^{\frac{\xi}{\sigma^2}} \cdot [\![(-1)\frac{\xi \bar{t}_n}{\sigma^2}]\!] = [\![\tilde{d}_n]\!]^{\frac{\xi}{\sigma^2}} \cdot [\![\bar{t}_n]\!]^{(-1)\frac{\xi}{\sigma^2}} \\
&= ([\![\tilde{d}_n]\!] \cdot [\![\bar{t}_n]\!]^{-1})^{\frac{\xi}{\sigma^2}} = [\![\frac{\tilde{d}_n}{\sigma^2} - \frac{\bar{t}_n}{\sigma^2}]\!]^{\xi} \\
&= [\![\frac{\tilde{d}_n - \bar{t}_n}{\sigma^2}]\!]^{\xi},
\end{aligned} \tag{8}$$

where $[\![\xi d_i]\!]$ denotes the element of scaled physiological case. Since ξ is large enough, $\frac{\xi}{\sigma^2}$ is guaranteed to be an integer. So $[\![\xi d_i]\!]$ is also an integer. Moreover, the scaled physiological case $[\![\xi \mathbf{d}]\!]$ can be written as

$$[\![\xi \mathbf{d}]\!] = ([\![\xi d_1]\!], [\![\xi d_2]\!], \cdots, [\![\xi d_n]\!]). \tag{9}$$

(5) After normalizing the PU's physiological case, HC will classify it.

4.4 Privacy-Preserving Physiological Case Classification

In the classification phase, PU only considers whether the physical condition is normal, so we presented the classification function of two-class problem involved SVM in this scheme. In addition, according to the description of Sect. 4.1, the kernel function in this scheme should be $(\mathbf{t}_i^T \cdot \mathbf{d})$. The specific classification process is as follows:

(1) By using the scaled physiological case $[\![\xi \mathbf{d}]\!] = ([\![\xi d_1]\!], [\![\xi d_2]\!], \cdots, [\![\xi d_n]\!])$ and normalized training case samples $\mathbf{t}_i = (t_{i1}, t_{i2}, \cdots, t_{in})$, HC computes the linear kernel in the form of ciphertext as

$$\begin{aligned} [\![\mathbf{k}_i]\!] &= [\![\xi d_1]\!]^{\xi t_{i,1}} \cdots [\![\xi d_n]\!]^{\xi t_{i,n}} \\ &= [\![\xi t_{i,1} \cdot \xi d_1]\!] \cdots [\![\xi t_{i,n} \cdot \xi d_n]\!] \\ &= [\![\xi t_{i,1} \cdot \xi d_1 + \cdots + \xi t_{i,n} \cdot \xi d_n]\!] \\ &= [\![\xi \mathbf{t}_i^T \cdot \xi \mathbf{d}]\!]. \end{aligned} \tag{10}$$

It is clearly shown that HC calculates the kernel function value without any interaction with PU.

(2) HC calculates the ciphertext value of the decision function as follows:

$$\begin{aligned} [\![d(\mathbf{d})]\!] &= [\![\xi^3 (\sum_{i \in S} \alpha_i y_i \mathbf{t}_i^T \mathbf{d} + b)]\!] \\ &= [\![\sum_{i \in S} \xi(\alpha_i y_i) \xi^2 (\mathbf{t}_i^T \cdot \mathbf{d}) + \xi^3 b]\!] \\ &= [\![\sum_{i \in S} \xi(\alpha_i y_i)(\xi \mathbf{t}_i^T \cdot \xi \mathbf{d}) + \xi^3 b]\!] \\ &= [\![\sum_{i \in S} \xi(\alpha_i y_i)(\xi \mathbf{t}_i^T \cdot \xi \mathbf{d})]\!] \cdot [\![\xi^3 b]\!] \\ &= [\![\xi^3 b]\!] \cdot \prod_{i \in S} [\![\mathbf{k}_i]\!]^{\xi(\alpha_i y_i)}. \end{aligned} \tag{11}$$

Note that, the $[\![d(\mathbf{d})]\!]$ cannot be sent directly to PU, because he/she can decrypt it using his/her private key to obtain the secret information of SVM (e.g., $\alpha_i, y_i, \mathbf{t}_i, b$, etc.). Therefore, we describe how to transmit the physiological case classification result to PU without disclosing the secret information of SVM in the next steps.

(3) Suppose $|d(\mathbf{d})| < 2^l$ and $l \in \mathbb{Z}_N$, HC can calculate the ciphertext of the classification label using integers A, B and $d(\mathbf{d})$ as

$$\begin{aligned} [\![cl]\!] &= [\![d(\mathbf{d})]\!]^A \cdot [\![B]\!] \\ &= [\![A \cdot d(\mathbf{d}) + B]\!], \end{aligned} \tag{12}$$

where $|A \cdot d(\mathbf{d}) + B| < 2^{\kappa'-2}$, integers A and B are used only once, and new A and B are generated in each initialization phase.

(4) Next HC sends the classification label $[\![cl]\!]$ to PU.

(5) Upon receiving $[\![cl]\!]$, PU can recover the classification label by using his/her private key and obtain the diagnosis result. The correctness of the diagnostic query is described as follows:

5 Security Analysis

In this section, we conduct a security analysis of the proposed EPPC scheme. In this regard, the analysis will focus on how the proposed EPPC scheme can realize the privacy preservation and the source authentication of the PU's physiological data and the confidentiality of the HC's SVM classification parameters.

Theorem 1. *(Privacy): The privacy of PU's physiological data is protected in the proposed EPPC scheme.*

`Proof`. In EPPC, the PU's physiological data $\tilde{\mathbf{d}} = (\tilde{d}_1, \tilde{d}_2, \cdots, \tilde{d}_n)$ is encrypted before it is sent to HC. Specifically, each element of $\tilde{\mathbf{d}}$ individually using OU homomorphic cryptosystem as $[\![\tilde{\mathbf{d}}]\!] = (g^{\tilde{d}_1} \cdot h^{\tilde{r}_1} mod N, g^{\tilde{d}_2} \cdot h^{\tilde{r}_2} mod N, \cdots, g^{\tilde{d}_n} \cdot h^{\tilde{r}_n} mod N) = ([\![\tilde{d}_1]\!], [\![\tilde{d}_2]\!], \cdots, [\![\tilde{d}_n]\!])$. Since OU cryptosystem is secure of indistinguishability under the condition of chosen-plaintext attack (IND-CPA [7], the physiological data $\tilde{\mathbf{d}} = (\tilde{d}_1, \tilde{d}_2, \cdots, \tilde{d}_n)$ is semantic secure and privacy-preserving. Therefore, even if the adversary breaks into the database of HC, he/she can only obtain the ciphertext of PU's physiological data instead of identifying its specific content. After receiving $[\![\tilde{\mathbf{d}}]\!] = ([\![\tilde{d}_1]\!], [\![\tilde{d}_2]\!], \cdots, [\![\tilde{d}_n]\!])$ from PU, HC normalizes $[\![\tilde{\mathbf{d}}]\!]$ as

$$\begin{aligned} [\![\xi\mathbf{d}]\!] &= ([\![\frac{\tilde{d}_1 - \bar{t}_1}{\sigma^2}]\!]^{\xi}, [\![\frac{\tilde{d}_1 - \bar{t}_2}{\sigma^2}]\!]^{\xi}, \cdots, [\![\frac{\tilde{d}_n - \bar{t}_n}{\sigma^2}]\!]^{\xi}) \\ &= ([\![\xi d_1]\!], [\![\xi d_2]\!], \cdots, [\![\xi d_n]\!]). \end{aligned} \tag{13}$$

However, since HC does not have the private key (p', q') and the calculation of each element in Eq. 13 is performed by HC without interacting PU, even if HC obtain the data, he still cannot identify the PU's physiological data $\tilde{\mathbf{d}} = (\tilde{d}_1, \tilde{d}_2, \cdots, \tilde{d}_n)$. Therefore, the privacy of PU's physiological data is protected in the proposed EPPC scheme.

Theorem 2. *(Authentication): The authentication of the PU's physiological case is fulfilled in the proposed EPPC scheme.*

Proof. The description of the proposed EPPC scheme shows that the PU's physiological case is signed by BLS short signature [1] as $\delta = x_u \cdot H([\![\tilde{\mathbf{d}}]\!]||Y||T)$. Because the BLS signature scheme is provably secure under the CDH problem in the random oracle model, the physiological case received by HC coming from a valid PU is assured. Meanwhile, the adversary $\mathcal{A}$'s malicious behaviors (such as falsifying signatures, modifying physiological case, etc.) can also be easily detected. Therefore, from the above analysis, the authentication of the PU's physiological case is achieved in the proposed EPPC scheme.

Theorem 3. *(Confidentiality): The training set and the SVM classification parameters of HC are confidential in the proposed EPPC scheme.*

Proof. In the privacy-preserving physiological case classification phase of the EPPC scheme, HC uses OU homomorphic cryptosystem to encrypt the linear kernel and the decision function, respectively. Specifically, the ciphertext of the linear kernel is $[\![\mathbf{k}_i]\!] = [\![\xi \mathbf{t}_i^T \cdot \xi \mathbf{d}]\!]$ and the ciphertext of the decision function is $[\![d(\mathbf{d})]\!] = [\![\xi^3 b]\!] \cdot \prod_{i \in S} [\![\mathbf{k}_i]\!]^{\xi(\alpha_i y_i)}$. Because it is very rare to obtain the private key (p', q'), even if the adversary $\mathcal{A}$ steals the decision function $[\![d(\mathbf{d})]\!]$ or linear kernel $[\![\mathbf{k}_i]\!]$, he is impossible to obtain the training set and the HC' SVM classification parameters.

Security analysis shows that our EPPC scheme can meet the security requirements.

6 Performance Evaluation

In this section, we will compare schemes EPPC and [10] from the following two aspects: computational cost and communication overhead.

6.1 Security Comparison

Firstly, we compare the security of scheme EPPC with that of related scheme [10]. The results of the comparison of performance characteristics are shown in Table 1. The two schemes all use homomorphic encryption scheme to encrypt data, and use SVM to obtain classification label. Therefore, patient privacy and SVM classification parameters can be protected in both schemes. However, the lack of message authentication in scheme [10] makes it possible for the patient's physiological data to be forged. In addition, scheme EPPC has obvious advantages in terms of the efficiency of obtaining diagnostic results. The specific analysis is as follows.

Table 1. Security comparison

Metrics/parameters	[10]	EPPC
Privacy preservation of patient	Yes	Yes
Privacy preservation of SVM classification parameters	Yes	Yes
Homomorphic encryption	Yes	Yes
SVM	Yes	Yes
Message authentication	No	Yes
Diagnostic efficiency	Low	High

6.2 Experimental Setup

All our evaluations were performed on Intel Core i7-6700 @3.10 GHz with 8 GB RAM. Our system ran Java with Win 10 64-bit. We performed our experiments using the code we wrote. The runtime of hash operation can be ignored because it is much smaller than the runtime of other operations. In addition, multiplication can be ignored compared to exponentiation in the same group operation. Here, we chose the same security parameters $|p'| = |q'| = 512$ bits to realize the same level of security.

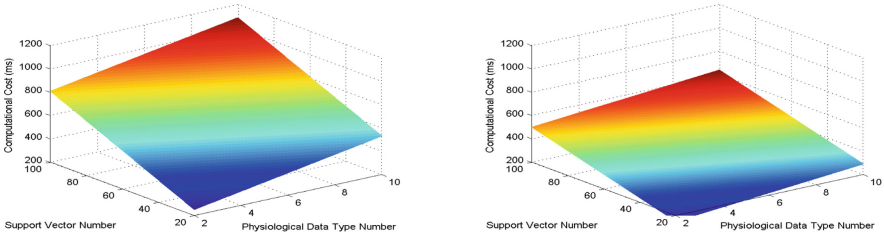
6.3 Computational Cost

In this subsection, we analyze the computational cost of PU and HC in each phase.

We will compare the computational costs with the scheme [10]. Table 2 presents the comparison results.

Table 2. Comparison of computational costs

Schemes	PU/Patient	HC/Server	Total
[10]	$2nT_{ep}$	$(3n + \|S\| + 6)T_{ep}$	$(5n + \|S\| + 6)T_{ep}$
EPPC	$2nT_{eou} + T_m$	$(2n + \|S\| + 1)T_{eou} + 2T_b$	$(4n + \|S\| + 1)T_{eou} + T_m + 2T_b$



(a) Cost for scheme [10] (b) Cost for EPPC

Fig. 2. Comparison of computational costs.

A comparison diagram of the total computational costs can be obtained from Table 2, as shown in Fig. 2. Figure 2 (a) and (b) represent the total computational cost of scheme [10] and EPPC, respectively. It is clearly shown that the EPPC scheme significantly decreases the computational cost for both patients and HC.

6.4 Communication Overhead

In this subsection, we analyse the communication overhead of the proposed EPPC scheme. It mainly focuses on the communication between PU and HC. First, we consider the PU-to-HC communication, in which PU generates his/her physiological case and sends this case to HC. Next, we consider the HC-to-PU communication, where HC generates the ciphertext of classification label and send it to PU.

Table 3 shows the communication overhead results of the two schemes.

Table 3. Comparison of communication overheads

Schemes	PU/Patient to HC/Server	HC/Server to PU/Patient	Total
[10]	$2048n$	2048	$2048(n+1)$
EPPC	$1536n + 260$	1536	$1536n + 1796$

7 Conclusion

In this paper, we have proposed an efficient and privacy-preserving physiological case classification scheme for e-healthcare system. This system employs the OU homomorphic cryptosystem, signature authentication technology and SVM algorithm to efficiently classify the physiological cases without leaking the privacy of PU and HC. The scheme can not only protect the physiological data of PU, but also prevent the leakage of HC classification parameters. What's important, the scheme has lower computational cost and communication overhead, so PU can quickly obtain diagnostic results in this scheme. Experimental results illustrate that the proposed EPPC scheme characterizes more efficient and practical.

Acknowledgment. We are grateful to the anonymous reviewers for their invaluable comments. Mingwu Zhang is the corresponding author. This research was supported by the National Natural Science Foundation of China (NSFC) under Grant No.61672010, the Ph.D research startup foundation of Hubei University of Technology under Grant No. BSQD2019023, and the open research project of The Hubei Key Laboratory of Intelligent Geo-Information Processing under Grant No. KLIGIP-2017A11.

References

1. Boneh, D., Lynn, B., Shacham, H.: Short signatures from the weil pairing. J. Cryptol. **17**(4), 297–319 (2004)

2. Byvatov, E., Schneider, G.: Support vector machine applications in bioinformatics. Appl. Bioinform. **2**(2), 67–77 (2003)
3. Kumar, P., Lee, H.J.: Security issues in healthcare applications using wireless medical sensor networks: a survey. Sensors **12**(1), 55–91 (2012)
4. Li, M., Lou, W., Ren, K.: Data security and privacy in wireless body area networks. IEEE Wirel. Commun. **17**(1), 51–58 (2010)
5. Liu, X., Deng, R., Choo, K, R., Yang, Y.: Privacy-preserving outsourced support vector machine design for secure drug discovery. IEEE Trans. Cloud Comput. **8**, 610–622 (2018). https://doi.org/10.1109/TCC.2018.2799219
6. Liu, X., Deng, R., Choo, K, R., Yang, Y.: Privacy-preserving reinforcement learning design for patient-centric dynamic treatment regimes. IEEE Trans. Emerg. Top. Comput. (2019). https://doi.org/10.1109/TETC.2019.2896325
7. Okamoto, T., Uchiyama, S.: A new public-key cryptosystem as secure as factoring. In: Proceedings of International Conference on the Theory and Applications of Cryptographic Techniques, pp. 308–318 (1998)
8. Punithavathi, P., Geetha, S., Karuppiah, M., Islam, S.H., Hassan, M.M., Choo, K.K.R.: A lightweight machine learning-based authentication framework for smart IoT devices. Inf. Sci. **484**, 255–268 (2019)
9. Rahulamathavan, Y., Phan, R.C.W., Veluru, S., Cumanan, K., Rajarajan, M.: Privacy-preserving multi-class support vector machine for outsourcing the data classification in cloud. IEEE Trans. Dependable Secure Comput. **11**(5), 467–479 (2014)
10. Rahulamathavan, Y., Veluru, S., Phan, R.C., Chambers, J.A., Rajarajan, M.: Privacy-preserving clinical decision support system using gaussian kernel based classification. IEEE J. Biomed. Health Inform. **18**(1), 56–66 (2014)
11. Tong, S., Koller, D.: Support vector machine active learning with applications to text classification. J. Mach. Learn. Res. **2**(1), 999–1006 (2002)
12. Vapnik, V.N.: An overview of statistical learning theory. IEEE Trans. Neural Netw. **10**(5), 988–999 (1999)
13. Wang, D., Cheng, H., He, D., Wang, P.: On the challenges in designing identity-based privacy-preserving authentication schemes for mobile devices. IEEE Syst. J. **12**(1), 916–925 (2018)
14. Wang, D., Li, W., Wang, P.: Measuring two-factor authentication schemes for real-time data access in industrial wireless sensor networks. IEEE Trans. Ind. Inform. **14**(9), 4081–4092 (2018)
15. Wang, G., Lu, R., Shao, J., Guan, Y.: Achieve privacy-preserving priority classification on patient health data in remote eHealthcare system. IEEE Access **7**(1), 33565–33576 (2019)
16. Xu, C., Wang, J., Zhu, L., Zhang, C., Sharif, K.: PPMR: a privacy-preserving online medical service recommendation scheme in ehealthcare system. IEEE Internet Things J. **6**(3), 5665–5673 (2019)
17. Yi, X., Bouguettaya, A., Georgakopoulos, D., Song, A., Willemson, J.: Privacy protection for wireless medical sensor data. IEEE Trans. Dependable Secure Comput. **13**(3), 369–380 (2016)
18. Zhang, L., Zhang, Y., Tang, S., Luo, H.: Privacy protection for e-health systems by means of dynamic authentication and three-factor key agreement. IEEE Trans. Ind. Electron. **65**(3), 2795–2805 (2018)
19. Zhang, Y., Lang, P., Zheng, D., Yang, M., Guo, R.: A secure and privacy-aware smart health system with secret key leakage resilience. Secur. Commun. Netw. **2018**(4), 1–13 (2018)