文章编号:1671-4598(2023)04-0219-06 DOI:10.16526/j.cnki.11—4762/tp.2023.04.034 中图分类号:TP309 文献标识码:A

面向医疗系统的隐私保护疾病预测研究

李 超[1],张艳玲[1],张清媛[2]
(1.山东第一医科大学 第二附属医院,山东 泰安 271000;
2.山东第一医科大学 医学信息工程学院,山东 泰安 271016)

摘要:为了提高医疗数据的隐私性并有效对疾病进行预测,针对从物联网(IoT)设备收集的患者医疗数据,构建了面向医疗数据的隐私保护疾病预测系统框架,通过加密组合文本建立密钥提高了系统认证阶段的隐私性,加强系统和信息传输的安全性;利用基于对数循环值的椭圆曲线密码体制(LR-ECC)提高了数据传输阶段的安全性,从而授权的医护人员可以在医院安全地下载患者数据;运用基于象群遗传算法的的深度学习神经网络(EHGA-DLNN)分类技术,在疾病预测系统(DPS)阶段,实现了疾病数据的有效分类预测;实验结果表明,LR-ECC方法在加密时间和解密时间效率方面高于其他加密方法,并且能够达到98.87%的安全级别,EHGA-DLNN方法在疾病预测分类准确率达到98.35%。

关键词:医疗系统;物联网;数据安全;隐私保护;疾病预测

Research on Privacy Protection and Disease Prediction for Medical System

LI Chao[1], ZHANG Yanling[1], ZHANG Qingyuan[2]
(1. Second Affiliated Hospital of Shandong First Medical University, Tai'an 271000, China;
2. School of medical information engineering, Shandong First Medical University, Tai'an 271016, China)

Abstract: In order to improve the privacy of medical data and effectively predict diseases, a privacy protection disease prediction system framework with facial medical data is constructed for patient medical data collected from Internet of Things (IoT) devices. The privacy of the system at the authentication stage is improved by encrypting the combination text to establish a key, and the security of the system and information transmission is strengthened. The elliptic curve cryptosystem based on logarithmic cyclic value (LR-ECC) is used to improve the security of the data transmission stage, so that authorized medical personnel can safely download patient data at the hospital. By using the deep learning neural network (EHGA-DLNN) classification technology based on image swarm genetic algorithm, the effective classification and prediction of disease data are realized in the stage of disease prediction system (DPS). The experimental results show that the LR-ECC method is superior to other encryption methods in terms of encryption time and decryption time efficiency, and can achieve a security level of 98.87%. The accuracy of the EHGA-DLNN method in disease prediction classification reaches 98.35%.

Keywords: medical system; internet of things; data security; privacy protection; disease prediction

0 引言

当前的医疗系统是一个复杂的数据驱动网络,依赖于对患者的连续监控、数据共享和流式传输[1],凭借先进的大数据分析为患者提供必要的医疗服务[2]。随着物联网(IoT)技术在现代医疗系统中的不断发展,为患者和医疗专业人员带来了便利。各种可穿戴式IoT监测设备采集不同类型的病理数据并最终传输到医院云服务器中,这对医疗系统的隐私性和安全性提出了更高的要求[3]。可穿戴式医疗数据采集设备有智能心电图机、蓝牙血糖测量设备和3G血压测量设备等,主要用于监测心率、血糖和血压等各种生理症状[4],并最终为疾病预测系统(DPS)提供数据支撑。然而,在现代医疗系统中,个人数据隐私越来越受到关注,尤其是涉及敏感的个人医疗数据。个人医疗数据面临非法共享机密信息、非法使用私人数据、个人身份、敏感数据暴露等系列隐私问题[5]。因此,医疗系统必须同时兼顾数据隐私的安全性和疾病预测的准确性。

对于患者数据隐私保护方面,文献[6]使用随机森林(RF)技术的隐私保护计算协议,结合属于多个终端的医疗系统提供安全训练,并能够做出准确疾病预测。此外,原始数据可执行安全和计算处理并保存在云中。文献[7]提出了隐私保护单决策树方法,使用同态加密方式保护用户数据,通过相同的密钥对生成随机数来防止攻击方解密数据,然而,同态加密提供的安全性较差。文献[8]开发了

收稿日期:2022-09-12; **修回日期**:2022-10-15。
基金项目:国家自然科学基金(81871356);山东省临床医学科技创新计划(202219044)。
作者简介:李 超(1973-),男,山东肥城人,大学本科,工程师,主要从事医疗信息化方向的研究。
张艳玲(1973-),女,山东泰安人,大学本科,副主任护师,主要从事公共卫生方向的研究。
张清媛(1968-),女,山东济南人,博士,教授,主要从事医疗信息化方向的研究。
引用格式:李 超,张艳玲,张清媛.面向医疗系统的隐私保护疾病预测研究[J].计算机测量与控制,2023,31(4):219-224,231.

一种基于深度 Q 学习的神经网络框架，采用隐私保护方法（DQ-NNPP），以保护从医疗物联网设备传输的敏感患者医疗数据免受外部威胁。所有这些模型中的数据保密性和安全性都较低。此外，在所有上述方法中也普遍缺乏预测效率和准确性。对于疾病预测方面，文献［9］提出了基于模糊神经分类器的医疗保健疾病预测系统，然而，云数据库中的医疗数据的安全性不足。文献［10］提出了基于多标签 k 邻近的多级医学预诊断系统（ML-kNN），通过服务器自动减少需要使用 k 均值聚类计算的医疗保健实例的数量，并结合基于 ML-kNN 分类向医疗保健用户提供服务。然而，该系统的计算复杂度较高。文献［11］提出了利用云和基于物联网的数据库的预测方法，用于预测使用从生物传感器收集的患者数据的疾病。

为了在保护患者数据安全的前提下对疾病进行有效预测，DPS 必须通过使用医疗数据建立分类预测模型，并且不能与第三方进行共享以防止数据泄密[12]。DPS 能够提高诊断的速度和准确性，并且提高医疗服务的质量，因此，预测效率也是构建 DSP 时应考虑的关键因素。文献［13］提出了一种有效且保密的疾病预测系统。患者就诊的医疗记录被加密并发送到云服务器，使用单层感知器学习方法训练预测模型，同时仍然保持患者隐私。文献［14］提出了以混合推理为中心的隐私感知疾病预测支持系统（PDPSS）。利用模糊集理论、k 邻近和以案例为中心的推理组合优势改进的预测结果，从而将疾病预测支持系统（DPSS）扩展到以 Pailliers 同态加密为中心的 PDPSS，从而保护患者的敏感数据免受非法用户访问，然而，该系统具有较高的通信和计算成本。

本文构建了隐私保护疾病预测系统架构，利用加密组合文本生成密钥提高了系统认证阶段的隐私性，提出了基于对数循环值的椭圆曲线密码体制（LR-ECC）的数据加密方法，增强了数据传输阶段的医疗数据的安全性。采用基于象群遗传算法的的深度学习神经网络（EHGA-DLNN）分类技术对 IoT 监测设备采集的感知数据进行分类，提高了 DPS 阶段疾病预测的准确性。

1 系统结构及原理

为了在医疗系统中有效地保护数据安全性并准确地预测疾病，本文构建了面向医疗系统的隐私保护疾病预测系统，实现了保护患者数据的隐私性和疾病预测的准确性。由于 IoT 传感器设备穿戴在患者身体上，并考虑到部署以太网的复杂性和 WIFI 网络的局限性，本文的整体网络架构采用窄带物联网（NB-IoT）技术原理，NB-IoT 是一种部署于 GMS 网络或 LTE 网络上的低频段、低速率的网络，支持低功率设备在广域网络的蜂窝数据连接，具有海量连接、深度覆盖、稳定可靠、综合成本低等优势。该系统分为系统认证、安全数据传输和疾病预测系统（DPS）3 个阶段。用户利用医院的移动 APP 或网站向相应医院进行注册，当使用有效的认证方法成功完成登录后，则传感器值将通过雾层安全上传至医院云服务器。同时，相应的医生可以安全地下载患者数据，并使用分类模型准确地完成疾病预测。本文提出的隐私保护疾病预测系统架构，如图 1 所示。

2 系统认证阶段

为了加强系统和信息传输的安全性，本文在医生、医护人员和医院云服务器、患者和医院云服务器以及医院和医院云服务器之间设计了认证功能。系统认证阶段分为注册、登录和验证 3 个部分。

2.1 注册

在各种 IoT 设备访问系统数据之前，需经过系统管理员的批准。当批准通过后，管理员将数据反馈给 IoT 设备进行认证。注册过程包括如下 4 个部分。

1）患者详细信息：

用户在注册时提供患者详细信息，患者详细信息包含用户名、患者姓名、性别、年龄、地址、密码、患者 ID、医院 ID、医生名称等，通过移动 APP 或网站输入并保存到数据库中。患者详细信息集可以表示为：

$$\widetilde{P}_{pd}=\{\widetilde{p}_i\},i\in[1,k] \tag{1}$$

其中：$\widetilde{p}_i$ 为患者信息，k 为患者信息的数量。

2）组合文本：

输入患者详细信息后，将用户 ID 和医院 ID 合并为组

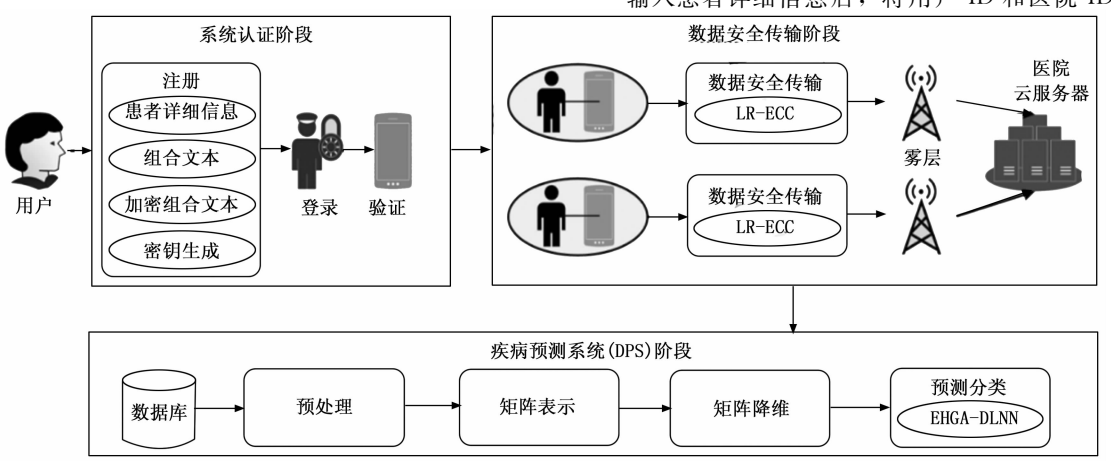

图 1 系统架构

合文本，组合文本可以表示为：

$$\vec{T}''_{ct} = \widetilde{p}_u \otimes \widetilde{p}_h \tag{2}$$

其中：$\widetilde{p}_u$ 和 $\widetilde{p}_h$ 分别为从 $\widetilde{P}_{pd}$ 中提取的用户 ID 和医院 ID，$\otimes$ 为合并函数。

3）加密组合文本：

在注册期间完成组合文本后执行加密过程，从而使用替换密码将组合文本转换为密码。替换密码是根据预设系统将明码转化为密码的加密方式。在所有的替换密码中，密码字母仅为明码字母的简单循环移位[15]。此外，密码元素包括数字、标点符号和 26 个字母。组合文本的加密可以表示为：

$$(\vec{T}''_{ct})_e = (\vec{T}''_{ct}) \bmod 26 \tag{3}$$

在验证时，密码由医院云服务器发送给数据所有者。当用户试图下载医院云服务器中任何文件时，则医院云服务器会要求用户输入密码。当用户输入并发送正确的密码时，医院云服务器将用户验证为授权用户，并允许用户访问数据。则替换密码可以表示为：

$$(\vec{T}''_{ct})_e \xrightarrow{\text{匹配}} (CS \xrightarrow{\text{确认}} \widetilde{A}_{au}) \tag{4}$$

$$(\vec{T}''_{ct})_e \xrightarrow{\text{不匹配}} (CS \xrightarrow{\text{确认}} \widetilde{A}_{ur}) \tag{5}$$

其中：$\widetilde{A}_{au}$ 和 $\widetilde{A}_{ur}$ 分别为授权用户和未授权用户，CS 为医院云服务器。

4）密钥生成：

医院云服务器负责创建公钥和私钥，其中，公钥自动生成并保存在医院云服务器中；私钥则发送到用户注册期间提供的电子邮件。公钥与私钥的数学表达式为：

$$CS \xrightarrow{(\vec{K}''_{pu}, \vec{K}''_{pr})} User \tag{6}$$

其中：$\vec{K}''_{pu}$ 为公钥，$\vec{K}''_{pr}$ 为私钥，$User$ 为用户。通过对 $\vec{K}''_{pu}$ 和 $\vec{K}''_{pr}$ 取整后计算对数值得到最终的密钥：

$$\vec{K}''_{se} = \log(\vec{K}''_{pu} \oplus \vec{K}''_{pr}) \tag{7}$$

其中：$\oplus$ 为 $\vec{K}''_{pu}$ 和 $\vec{K}''_{pr}$ 的取整。

2.2　登录

登录功能是用于验证用户的凭证集。当用户请求登录时，则医院云服务器负责验证用户输入的用户 ID 和密码。如果用户提供了正确的私钥，则通过医院云服务器执行文本文件的解密并通过雾层传输给用户。当私钥不正确时，则拒绝用户访问医院云服务器。

2.3　验证

验证功能在系统登录后执行。系统将匹配用户 ID 和密码。如果所有详细信息都匹配，则系统最终确定用户已向相应的 HCS 完成注册。否则，系统返回到注册阶段。

3　数据安全传输阶段

本文利用对数循环值的椭圆曲线密码体制（LR-ECC）[16]对 IoT 设备采集的感知数据进行加密，并通过雾层发送到医院云服务器。椭圆曲线密码（ECC）[17]是以密钥为中心的数据加密方法，通过公钥和私钥对 web 流量进行解密和加密。用户在成功登录后，利用发送方的私钥 $\vec{K}''_{pr}$、接收方的公钥 $\vec{K}''_{pu}$ 和生成密码的密钥 $\vec{K}''_{se}$ 对文本 $\widetilde{C}_{tx}$ 进行加密，加密文本可以表示为：

$$E(\widetilde{C}_{tx}) = \widetilde{C}_{tx} + (Rn \cdot \vec{K}''_{pr}\vec{K}''_{pu}) \cdot \vec{K}''_{se} \tag{8}$$

其中：Rn 为$(1, n-1)$范围内的随机数。加密文本 $E(\widetilde{C}_{tx})$ 通过医院云服务器发送并反向执行 LR-ECC 解密文本 $E(\widetilde{C}_{tx})$，从而最终在用户终端恢复为文本 $\widetilde{C}_{tx}$。

4　疾病预测系统（DPS）阶段

DPS 是预测疾病在患者身上出现可能性的过程，利用 IoT 感测值来确定患者是否患有疾病。DPS 阶段分为数据收集、预处理、矩阵表示、矩阵降维和分类 5 个部分。

4.1　数据收集

DPS 收集疾病数据集并形成数据库，用于处理的疾病数据集可以表示为：

$$H''_{ds} = \{h_j\}, j \in [1, m] \tag{9}$$

其中：h_j 为疾病数据信息，m 为疾病数据的数量。

4.2　预处理

由于 IoT 采集的数据集可能包含重复数据，为了避免重复训练相同的信息，降低系统计算的复杂度。本文通过重复数据删除方式[18]来消除冗余数据。重复数据删除是通过避免重复数据存储和节省带宽来有效管理云存储空间的方法。重复数据删除分为缺失值插补和最小最大归一化两个阶段：

1）缺失值插补。当 IoT 采集的数据集出现缺失值时，采用缺失值插补处理数据；将缺失数据转换为替代值的方法称为插补[19]。任何记录都包含少量缺失值，但可以通过改变针对特定属性的缺失值与平均值来加载生成缺失值，从而完成缺失值插补。

2）最小最大归一化。当 IoT 采集的数据集消除数据冗余（或重复）和不需要的特征（即插入、更新和删除异常）时，采用最小最大归一化处理数据。最小最大归一化是通过改变特定范围内的数据值，在系统中实现归一化的同时产生高效的输出[20]。利用最小值和最大值在 0～1 之间或在 −1～1 之间更改数据值。通过对每个数据分别进行最小最大处理得到归一化值：

$$MinMax = \frac{H''_{ds} - min(H''_{ds})}{max(H''_{ds}) - min(H''_{ds})} \tag{10}$$

以最小值和最大值为中心对缺失值进行替换，可以有效地提高数据的完整性。

4.3　矩阵表示

将预处理后的数据作为矩阵表示。通常，预处理后的数据表示为 $L \times P$ 矩阵，其中，L 为实例数，P 为实验中存在的属性数，即性别、年龄、地址等。矩阵中的每个单元格 $I_{i,j}$ 均可等效，则预处理后的数据矩阵表示为：

$$\boldsymbol{M}''_{\text{rep}} = \begin{bmatrix} I_{1,1} & I_{1,2} & \cdots & I_{1,n} \\ I_{2,1} & I_{2,2} & \cdots & I_{2,n} \\ \cdots & \cdots & \cdots & \cdots \\ I_{n,1} & I_{n,2} & \cdots & I_{n,n} \end{bmatrix} \tag{11}$$

4.4 矩阵降维

矩阵降维是将高维空间转换为低维空间的数据转换，从而使低维数据保留原始数据的关键属性。本文利用高斯核函数线性判别分析（GK-LDA）算法[21]对预处理后的矩阵表示进行降维。其中，线性判别分析（LDA）通过检测方向来降低类间散度与类内散度比来加密信息，并在降维过程中实现监督功能[22]。高斯核函数（GK）用于提高降维后的数据还原精度[23]。GK-LDA 算法的具体步骤如下。

步骤 1：输入预处理后的数据矩阵表示 $\boldsymbol{M}''_{\text{rep}}$。

步骤 2：利用 $\widetilde{\boldsymbol{B}}_c$ 和 $\widetilde{\boldsymbol{W}}_c$ 分别表示类间和类内的散度矩阵：

$$\widetilde{\boldsymbol{B}}_c = \sum_{m=1}^{a} s_m((\boldsymbol{M}''_{\text{rep}})_m - \boldsymbol{M}''_{\text{rep}})((\boldsymbol{M}''_{\text{rep}})_m - \boldsymbol{M}''_{\text{rep}})^T \tag{12}$$

$$\widetilde{\boldsymbol{W}}_c = \sum_{m=1}^{a}\sum_{n=1}^{q_m}(D_n - (\boldsymbol{M}''_{\text{rep}})_m)(D_n - (\boldsymbol{M}''_{\text{rep}})_m)^T \tag{13}$$

其中：

$$(\boldsymbol{M}''_{\text{rep}})_m = \frac{1}{s_m}\sum_{D_n \in D_m} D_n \tag{14}$$

$$\boldsymbol{M}''_{\text{rep}} = \frac{1}{s}\sum_{m=1}^{s}\sum_{D_n \in D_m} D_n \tag{15}$$

其中：D_n 和 D_m 为数据点，s_m 为第 m 类样本的数量，s 为数据矩阵表示中样本的数量。

步骤 3：为了提高降维后的数据还原精度，利用 GK 计算数据点之间的距离，即权重：

$$\kappa(D_m, D_n) = \exp(-\gamma_{mn}\| D_m - D_n \|^2) \tag{16}$$

其中：γ_{mn} 为数据点 D_n 和 D_m 之间的超参数。

步骤 4：在 LDA 搜索线性子空间 $R(c-1$ 分量）中，利用不同类别的投影得到最佳划分，结合最大化后续判别准则可得：

$$V(R) = max\ \frac{Tom(\boldsymbol{R}^T\widetilde{B}_c\boldsymbol{R})}{Tom(\boldsymbol{R}^T\widetilde{W}_c\boldsymbol{R})} \tag{17}$$

其中：$Tom(\cdot)$ 为矩阵的迹，$\boldsymbol{R}$ 为特征向量。除了 $\boldsymbol{R}$ 的正交约束外，可以计算广义特征向量和特征值：

$$\widetilde{B}_c\boldsymbol{R}_m = \lambda_m\widetilde{W}_c\boldsymbol{R}_m \tag{18}$$

其中：$\boldsymbol{R}_m$ 和 λ_m 分别为 $\widetilde{B}_c$ 关于 $\widetilde{W}_c$ 的第 m 个广义特征向量和特征值。

步骤 5：通过减少特征值对特征向量进行排序，将预处理后的数据矩阵表示 $\boldsymbol{M}''_{\text{rep}}$ 的整个输入矩阵描述的线性组合生成的矩阵降维集为：

$$(\boldsymbol{H}_{\text{rf}})_v = (\boldsymbol{M}''_{\text{rep}})_m \cdot R_m \tag{19}$$

4.5 基于象群遗传算法的的深度学习神经网络（EHGA-DLNN）分类

将 $(\boldsymbol{H}_{\text{rf}})_v$ 输入到分类器，利用 EHGA-DLNN 算法[24]对训练数据进行分类。正常的深度学习神经网络（DLNN）[25]也可对数据进行分类，然而，随机权重值对正常和疾病严重程度的分类精度较低。利用遗传算法（GA）[26]对象群进行优化，可以减少 DLNN 算法中的反向传播问题。DLNN 分为输入层、隐藏层和输出层。

4.5.1 输入层

将预处理后的矩阵降维集 $(\boldsymbol{H}_{\text{rf}})_v$ 用于训练系统，并确定其等效权重，其描述如下：

$$(\boldsymbol{H}_{\text{rf}})_v = \{h_i\}, i \in [1,k] \tag{20}$$

$$(\boldsymbol{W}_{\text{ew}})_v = \{w_i\}, i \in [1,k] \tag{21}$$

象群优化算法（EHO）是以种群为中心的优化技术，主要分为氏族更新算子和分离算子[27]。将 GA 的交叉变异与 EHO 中的更新步骤相结合得到象群遗传算法（EHGA）[28]，从而提高搜索精度。EHGA 具体步骤如下：

初始化种群空间、置信空间和可调算子。当涉及到氏族更新时，通过搜索策略更新大象的位置：

$$T^k_{n,xl} = T^k_{xl} + \alpha(T_{b,xl} - T^k_{xl})g \tag{22}$$

其中：$T^k_{n,xl}$ 和 T^k_{xl} 分别为大象 k 在氏族 xl 上的新位置和旧位置，$T_{b,xl}$ 为氏族 xl 的族长，g 为 [0，1] 范围内的随机数，α 为族长效应的比例因子。在氏族更新过程之后，氏族中最差的大象被淘汰，并且在搜索空间中使用分离算子任意生成新位置：

$$T_{w,xl} = T[(T_{\min})_{\max} \cdot f_r]_{\min} \tag{23}$$

其中：$T_{w,xl}$ 为氏族 xl 上具有最差适应值的公象，$T_{\min}$ 和 $T_{\max}$ 分别为搜索空间的上限和下限，f_r 为 [0，1] 范围内的随机数。

将 GA 的交叉变异与 EHO 中的更新步骤相结合，从而提高搜索精度。因此，在公式（23）更新新位置之前，利用交叉变异中的两点交叉可以得到：

$$T(t+1) = T_{f_r}(t) + C_1 + C_2 \tag{24}$$

$$C_1 = \frac{|\ T(t)\ |}{3} \tag{25}$$

$$C_2 = C_1 + \frac{|\ T(t)\ |}{2} \tag{26}$$

其中：t 为迭代次数，C_1 和 C_2 为交叉点的两个点。

使用 EHO 算法更新新位置得到最优权重：

$$(O'_w)_v = T(t+1) \cdot T_{b,xl} \tag{27}$$

将输入值乘以权重后累加得到分配值：

$$\tilde{e}_v = \sum_{v=1}^{n}((H_{rf})_v \cdot (O'_w)_v) \tag{28}$$

因此，输入到隐藏层的激活函数为：

$$\vec{A}_v = f\Big(\sum_{v=1}^{n}((H_{rf})_v \cdot (O'_w)_v)\Big) \tag{29}$$

4.5.2 隐藏层

隐藏层的输出值为：

$$\vec{H}_v = Bias + \sum_{v=1}^{n}\vec{A}_v \cdot (O''_w)_v \tag{30}$$

其中：$Bias$ 为偏置值。

4.5.3 输出层

输出层的输出值为：

$$\vec{O}_v = Bias + \sum_{v=1}^{n}\vec{H}_v \cdot (O''_w)_v \tag{31}$$

计算损失函数为：

$$L_v = \vec{G}_v + \vec{O}_v \tag{32}$$

其中：$\vec{G}_v$ 为神经网络的期望结果。在这种情况下，损失函数的阈值设置为最小值。如果初始化的阈值满足此适应

度，则输出为最终输出。如果初始化的阈值与该适应度不匹配，则重新更新随机权重值，并且相同的 EHGA 优化随机权重值。使用该 EHGA-DLNN 算法再次确定输出单元，并训练输出数据来进行分类。同时，EHGA-DLNN 算法继承了 GA 算法，保留了 GA 算法快速收敛的特点，解决了 DLNN 算法反向传播过程带来的梯度消失问题。EHGA-DLNN 算法的伪代码，如算法 1 所示。

算法 1：EHGA-DLNN 算法输入：矩阵降维集 $(\boldsymbol{H}_{rf})_v$

输出：受疾病影响的分类数据

初始化 $(W_{ew})_v$、$Bias$、$\vec{A}_v$ 和 $\vec{H}_v$

计算训练样本数 λ

if $\lambda = 0$

错误（λ 不是整数）

end if

for 每个减少的数据 do

使用 EHGA 更新权重值的位置

使用如下公式更新新位置

$(O'_w)_v = T(t+1) \cdot T_{b,st}$

while v 小于迭代次数 do

则使用如下公式执行激活函数

$$\vec{A}_v = f\left(\sum_{v=1}^{n}((H_{rf})_v \cdot (O'_w)_v)\right)$$ //计算激活函数

for $\vec{H}_v$ do

计算隐藏层的输出

$$\vec{H}_v = Bias + \sum_{v=1}^{n}\vec{A}_v \cdot (O'_w)_v$$

计算输出层的输出

$$\vec{O}_v = Bias + \sum_{v=1}^{n}\vec{H}_v \cdot (O'_w)_v$$

end for

end while

5　实验分析

为了对患者数据在传输过程实现隐私保护，并将 IoT 感测数据按严重程度分为正常数据和疾病数据。利用医疗数据集在 JAVA 中运行实验评估本文方法的有效性。在数据安全传输阶段，利用数据安全传输时间和安全级别分析所提出的 LR-ECC 方法的安全性。在 DPS 阶段，采用不同的分类算法与所提出的 EHGA-DLNN 方法对比验证疾病分类预测有效性。EHGA-DLNN 算法的具体参数设置如下：1）EHO 中的象群种群数量为 5，最大迭代次数为 100，族长效应比例因子为 0.5；2）GA 中的交叉率设置为 0.6，变异率为 0.2；3）DLNN 中的学习率为 0.01，衰减系数为 0.99。

5.1　评估参数

本文选取加密时间、解密时间、准确率、灵敏度、特异性、精度、召回率和 F1 值等 8 个指标作为评估参数。

1）加密时间：加密开始和结束时间与加密算法从明码构造密码所用时间之差。

2）解密时间：加密开始和结束时间之差。

3）准确率：准确地确定记录可能是正常或受疾病影响的概率。

4）敏感性：正确区分正常和疾病的比率。

5）特异性：影响总分类结果的疾病分类准确率。

6）精度：对于某个类别，准确记录属于该类别的概率。

7）召回率：对于特定类别，数据集中所有可用记录中准确预测受疾病影响结果的计数。

8）F1 值：利用精度和召回率对模型进行整体估计的调和均值。

5.2　安全级别性能分析

本文利用加密时间和解密时间分析所提出 LR-ECC 方法在数据安全传输的性能，并与与椭圆曲线密码编码学（ECC）[17]、非对称加密（RSA）[29]、全同态加密（FHE）[30]和迪菲－赫尔曼（DH）算法[31]进行比较。数据安全传输时间性能比较，如图 2 所示。

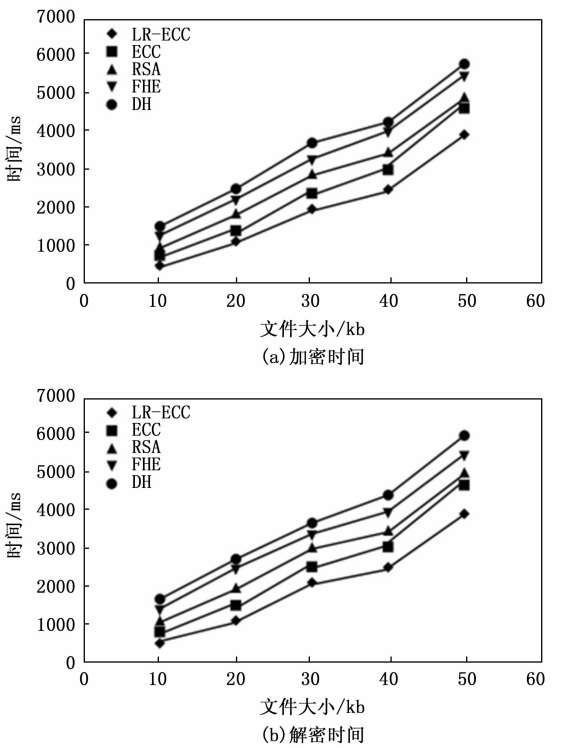

图 2　数据安全传输时间性能对比

由图 2 可见，为了安全传输 10～50 kB 不等的文件。对于加密时间指标，本文提出的 LR-ECC 方法加密 10 kB 的文件仅需 465 ms。相比之下，ECC、RSA、FHE 和 DH 算法分别需要 801 ms、1 013 ms、1 346 ms 和 1 646 ms 执行加密。类似地，LR-ECC 方法对 20～50 kB 的文件加密实现了更优异的性能。对于解密时间指标，LR-ECC 方法解密 10 kB 的文件仅需 475 ms。相比之下，其他方法的性能明显低于 LR-ECC 方法。类似地，LR-ECC 方法对 20～50 kB 的文件解密同样实现了更优异的性能。因此，在数据安全传输阶段，与其他方法相比，本文提出的 LR-ECC 方法在传输时间效率方面具有较高的性能。

本文提出的 LR-ECC 方法与 ECC、RSA、FHE 和 DH 算法的安全级别对比，如图 3 所示。

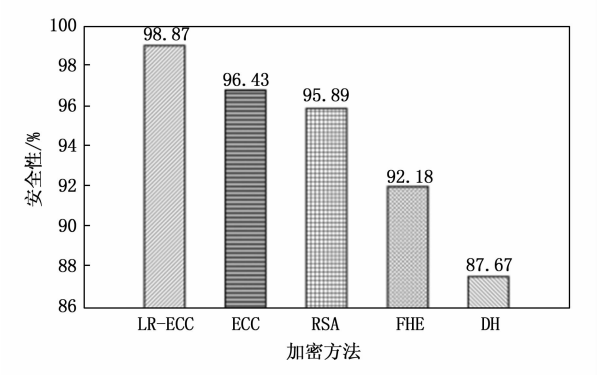

图 3 安全级别对比

由图 3 可见，本文提出的 LR-ECC 方法安全性达到 98.87%，而 ECC、RSA、FHE 和 DH 算法的安全性分别为 96.43%、95.89%、92.18%和 87.67%。因此，在数据安全传输阶段，与其他方法相比，本文提出的 LR-ECC 方法在传输安全性方面具有较高的性能。

5.3 疾病预测分类性能分析

利用人工神经网络（ANN）[32]、DLNN[25]、kNN[10]、支持向量机（SVM）[33]与本文提出的 EHGA-DLNN 方法对疾病预测分类进行对比。如表 1 所示。

表 1 分类方法性能比较

性能指标	分类方法				
	EHGA-DLNN	DLNN	ANN	kNN	SVM
准确率	98.35	95.33	93.35	92.33	91.23
灵敏度	97.33	95.56	92.32	90.45	89.33
特异性	96.36	94.57	89.99	88.13	86.33
精度	95.32	93.46	92.37	90.23	89.69
召回率	96.69	94.59	93.75	92.35	91.87
F1 值	96.37	94.57	93.35	92.97	91.12

由表 1 可见，与其他分类方法相比，所提出的 EHGA-DLNN 方法在疾病预测分类方面具有较高的性能。同时，就 F1 值和召回率等统计量而言，取得的结果表明，在 DPS 阶段，与其他方法相比，本文提出的 EHGA-DLNN 方法能够更快、更准确地预测疾病的严重程度。

6 结束语

在使用 IoT 设备的现代医疗系统中，患者的隐私和敏感医疗数据的安全性面临风险。同时，为了有效地通过 IoT 感测数据及时对疾病预测进行分类，本文构建了面向医疗系统的隐私保护疾病预测系统，实现了保护患者数据的隐私性和疾病预测的准确性。通过加密组合文本方式建立了系统加密认证，采用 LR-ECC 算法对 IoT 设备采集的感知数据进行加密，结合 EHGA-DLNN 分类技术对疾病数据进行分类。所提出的 LR-ECC 方法在加密时间、解密时间和安全级别分析方面，均优于传统 ECC、RSA、FHE 和 DH 算法的性能，且能够达到 98.87%的安全级别。同时，EHGA-DLNN 方法在疾病预测分类方面，均优于现有 DLNN、ANN、KNN 和 SVM 的性能，且准确率达到 98.35%。实验结果表明，本文方法在医疗系统的安全性和疾病预测的准确性均具有更高的效率。在未来的研究中，可使用更通用的策略来改进模型，从而在保持更高的安全性和隐私性的同时处理其他数据集类型。

参考文献：

[1] 陈庆龙，石春花，郝文延. 物联网医疗系统安全和隐私保护方法研究 [J]. 医学信息学杂志，2022，43 (1)：67－72.
[2] 刘 利，倪 丽，王 霞，等. 移动医疗设备信息隐私保护的雾计算解决方案研究 [J]. 中国医学装备，2022，19 (1)：138－142.
[3] 丁 淼，黄 晞. 基于云计算的医疗护理隐私信息部分加密系统设计 [J]. 微型电脑应用，2021，37 (11)：26－28，31.
[4] 薛 明. 基于超声图像的 PICC 智能穿刺医疗机器人控制系统设计 [J]. 计算机测量与控制，2020，28 (10)：86－90，95.
[5] 张晓敏. 大数据加密算法在数据安全保护中的应用研究 [J]. 计算机测量与控制，2021，29 (5)：204－208.
[6] ZMAB C，JMA C，YMAB C，et al. Privacy-preserving and high-accurate outsourced disease predictor on random forest [J]. Information Sciences，2019，496 (1)：225－241.
[7] 杨倩倩，王 龙，张晓娜. 基于数据挖掘的移动互联网数据包安全检测技术分析 [J]. 电子技术与软件工程，2022 (11)：18－21.
[8] KATHAMUTHU N D，CHINNAMUTHU A，IRUTHAYANATHAN N，et al. Deep Q-learning-based neural network with privacy preservation method for secure data transmission in internet of things (IoT) healthcare application [J]. Electronics，2022，144 (11)：157－168.
[9] 王 星，刘晓燕. 医疗大数据环境下的疾病预测模型研究 [J]. 制造业自动化，2022，44 (7)：24－27.
[10] 郭 凯，艾菊梅. 基于 FLANN 改进的 KNN 医疗分类算法 [J]. 计算机与现代化，2022 (8)：25－29.
[11] 张宇峰，乞国钰，周耀鉴，等. 基于人工智能及物联网下智慧养老系统的研究 [J]. 电子技术与软件工程，2021 (13)：206－207.
[12] 刘 焱，付剑锋，胡家锴. 基于分类模型的心血管疾病预测研究 [J]. 电脑编程技巧与维护，2021 (11)：23－24，30.
[13] 叶 琳，石胜源，罗铁清. AdaBoost 算法在乳腺癌疾病预测中的研究 [J]. 计算机时代，2021 (7)：61－64.
[14] 庞维庆，何 宁，罗燕华，等. 基于数据融合的 ABC－SVM 社区疾病预测方法 [J]. 浙江大学学报（工学版），2021，55 (7)：1253－1260.
[15] 杜成龙. 基于高安全等级的云存储信息架构及设计分析 [J]. 信息技术，2019，43 (11)：160－163.
[16] 李 伟，曾 涵，陈 韬，等. 基于动态补偿的椭圆曲线密码低成本抗功耗攻击策略及硬件结构研究 [J]. 电子与信息学报，2021，43 (9)：2439－2448.
[17] 李昊远，匡晓云，杨祎巍，等. 密码芯片中椭圆曲线算法的安全防护研究 [J]. 集成电路应用，2021，38 (4)：9－11.
[18] 宋桂平. 重复数据删除技术在云存储中的应用 [J]. 科技创新与应用，2022，12 (19)：158－161.

（下转第 231 页）