Future Generation Computer Systems 130 (2022) 140–154

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Identifying fraud in medical insurance based on blockchain and deep learning

Guoming Zhang [a], Xuyun Zhang [b], Muhammad Bilal [c,*], Wanchun Dou [a,*], Xiaolong Xu [d], Joel J.P.C. Rodrigues [e,f]

[a] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, 210023, China
[b] Department of Computing, Faculty of Science and Engineering Macquarie University, Sydney, NSW 2109, Australia
[c] Department of Computer and Electronics Systems Engineering, Hankuk University of Foreign Studies, Yongin-si, Gyeonggi-do, 17035, Republic of Korea
[d] School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, 210044, China
[e] Senac Faculty of Ceará, Fortaleza-CE, 60160-194, Brazil
[f] Instituto de Telecomunicações, Covilhã, Portugal

ARTICLE INFO

Article history:
Received 26 May 2021
Received in revised form 26 September 2021
Accepted 10 December 2021
Available online 16 December 2021

Keywords:
Medical big data
Anti-fraud
Blockchain
Deep learning

ABSTRACT

With the rapid growth of medical costs, the control of medical expenses has been becoming an important task of Health Insurance Department. Traditional medical insurance settlement is paid on a per-service basis, which leads to lots of unreasonable expenses. To cope with this problem, the single-disease payment mechanism has been widely used in recent years. However, the single-disease payment also has a risk of fraud. In this work, we propose a framework to identify fraud of medical insurance based on consortium blockchain and deep learning, which can recognize suspicious medical records automatically to ensure valid implementation on single-disease payment and lighten the work of medical insurance auditors. An explainable model BERT-LE is designed to evaluate the reasonability of ICD disease code for Medicare reimbursement by predicting the probability of a disease according to the chief complaint of a patient. We also put forward a storage and management process of medical records based on consortium blockchain to ensure the security, immutability, traceability, and auditability of the data. The experiments on two real datasets from two 3A hospitals demonstrate that the proposed solution can identify fraud effectively and greatly improve the efficiency in medical insurance reviews.

1. Introduction

With the rapid development of medical informatization, the information systems of hospitals have accumulated a large amount of data, which makes the medical industry enter the era of big data. Medical big data has brought tremendous value to medical field, and attracted lots of attention from both academia and industry [1]. The control of medical expense is an important research branch of medical big data.

In the traditional medical insurance system, medical expenses are paid on the basis of medical service items, which leads to excessive medical treatment and rising of medical expenses. To cope with the problems in item-based charge mechanism, the single-disease payment based on Diagnosis-Related Groups (DRGs) has been extensively studied and applied [2]. In the single-disease payment mechanism, a fixed payment standard is determined for each disease [3]. The social medical insurance agency pays the patients' hospitalization fees to hospitals according to the prescribed standard of each disease. To better understand the difference between single-disease payment mechanism and item-based payment mechanism, let us see an example. Suppose that a patient's diagnosis is pneumonia, according to the traditional medical insurance settlement mechanism, if the expenses covered by the medical insurance are 100,000 yuan and the reimbursement ratio is 90%, then the reimbursement expenses will be 90,000 yuan. In the DRGs payment model, assume that the DRG weighting rate of pneumonia is 6.2 and the unit payment price is 14,000 yuan, then the medical insurance reimbursement will be $6.2 \times 14{,}000 = 86{,}800$ yuan. After applying single-disease payment mechanism, the income of medical institutions is only related to each medical case and diagnosis and has nothing to do with the actual cost of treating a patient. If the treatment cost of a disease exceeds the standard of its group, the hospital need pay for the extra cost, so that the medical insurance expenditure can be controlled more effectively. It can standardize the

* Corresponding authors.
E-mail addresses: m.bilal@ieee.org (M. Bilal), douwc@nju.edu.cn (W. Dou).

https://doi.org/10.1016/j.future.2021.12.006
0167-739X/

medical resource utilization, that is, the consumption of medical institutions is directly proportional to the number of inpatients, the complexity of diseases, and the intensity of services. In short, the single-disease payment model stipulates a fixed medical insurance payment standard for each disease to avoid excessive medical behavior, and thereby decreases medical expenses. This mechanism guarantees the quality of medical services and is easy to operate.

However, the single-disease payment model may cause medical insurance fraud. For instance, when assigning the diagnostic code for an inpatient, the health care provider may change the actual low-cost disease code to another high-cost disease code to gain more revenue from medical insurance agency. Due to the large number of inpatients, the cost of reviewing each inpatient's medical records manually is extremely high. Take Jiangsu province of China for example, the number of inpatients each year is about 15 million, so it is impossible to audit each discharge diagnosis manually. Therefore, how to efficiently detect possible fraud has become an urgent problem in the single-disease payment mechanism. In the light of the above problem, we propose a framework to identify fraud of medical insurance based on blockchain and deep learning. In the framework, we transform the fraud identification problem into a text classification problem, and design an explainable model BERT-LE to estimate the reasonability of disease diagnosis by predicting the probability of diagnostic code according to inpatients' chief complaint. Only the identified abnormal diagnostic codes need to be manually audited, which can significantly improve the audit efficiency and reduce the workload of insurance auditing. In addition, in order to preserve the evidence of medical insurance fraud, we put forward a protocol of medical data storage and management based on consortium blockchain. The medical records are digitally signed and stored in the blockchain, which ensures the security and tamper-proofing of data, and at the same time enables the traceability and non-repudiation of fraudulent behavior. We also set up a credit rating mechanism for doctors and hospitals based on blockchain to carry out effective medical insurance supervision.

To validate our solution, we carry out experiments on two real datasets which contains 620,000 inpatients' medical records from two large hospitals. The results reveal that our model works well for medical insurance anti-fraud and has good explainability. Our proposed model can also help health departments evaluate the quality of medical records and aid doctors in giving accurate diagnostic codes.

The contributions of this paper are summarized as follows:

1. We propose a framework to transform the medical insurance anti-fraud problem into a text classification problem. The framework can effectively identify medical insurance fraud and reduce the workload of auditors.
2. We design a label embedding method to explain the logic of classification decisions, which can help users better understand and trust our model.
3. We put forward a protocol of medical data storage and access based on consortium blockchain, which can prevent the illegal tampering of medical records and make medical records traceable and auditable.
4. We conduct lots of experiments on two real datasets from large hospitals, and the results show that our model can play an effective role in anti-fraud for medical insurance and have good explainability.

The rest of this paper is organized as follows. In Section 2, the preliminaries and problem statement are presented. In Section 3, an explainable deep learning-based method to identify fraud of medical insurance is described. In Section 4, the blockchain-based medical records management process is elaborated. In Section 5, the proposed framework is evaluated and discussed. In Section 6, previous studies are reviewed. Finally, Section 7 concludes the paper and provides some future work discussions.

A preliminary version of this work has been reported in a conference short paper [4].

2. Preliminaries

2.1. Notations

The key medical term definitions in the proposed framework are provided as follows.

Definition 1. Chief complaint [5]. In medical domain, a chief complaint is a patient's description of his/her symptoms or (and) conditions, and duration of problems, etc., written by the physician into the patient's medical record. The chief complaint is the first item in the hospital medical record. It must reflect the characteristics of the first diagnosis of disease, and its description must be concise, refined and accurate, generally no more than 20 Chinese characters in the writing criteria of Chinese electronic medical record (EMR).

Definition 2. International Classification of Diseases (ICD) [6]. ICD is an internationally unified disease classification standard, represented by coding method to classify diseases according to pathology, etiology, anatomical location, and clinical manifestations. It is developed by the World Health Organization, and is widely used in the field of healthcare, such as determination of cause of death, disease statistics and reporting, evaluation of medical quality, and medical reimbursement. The 10th revision of this standard, known as "International Statistical Classification of Diseases and Related Health Problems", retains the abbreviation of ICD and is collectively referred to as ICD-10, which is widely used in Chinese medical institutions.

2.2. Problem statement

Our goal is to detect medical insurance fraud by recognizing the unreasonable ICD codes. The overall anti-fraud framework is shown in Fig. 1.

The data of doctors interacting with patients (including consultation, treatment process, etc.) are recorded and stored in the hospital EMR system. The chief complaint of the patient needs to be uploaded to the medical consortium blockchain after admission. When the patient discharges from hospital, the ICD code of his/her disease diagnosis and other related medical records are uploaded to the consortium blockchain. When reimbursing medical expenses, the medical records will be sent to medical insurance anti-fraud system, which will evaluate the reasonability of the ICD code. The payment of medical insurance depends on whether the ICD code is reasonable. If the ICD code is reasonable, the expense will be paid directly, otherwise, the medical records will be manually reviewed.

3. Identify fraud of medical insurance

To identify abnormal medical records, the reasonability of the disease diagnosis needs to be evaluated based on patients' chief complaints. For this purpose, we transform the identification of fraudulent medical records into a text classification task, which is to predict the probability of each ICD-10 code according to a chief complaint. Then the predicted probabilities of all ICD-10 codes will be sorted in descending order. If the ICD-10 code assigned to the medical record is in the top-k set of the predicted results, the medical record will be considered reasonable. Otherwise, it will

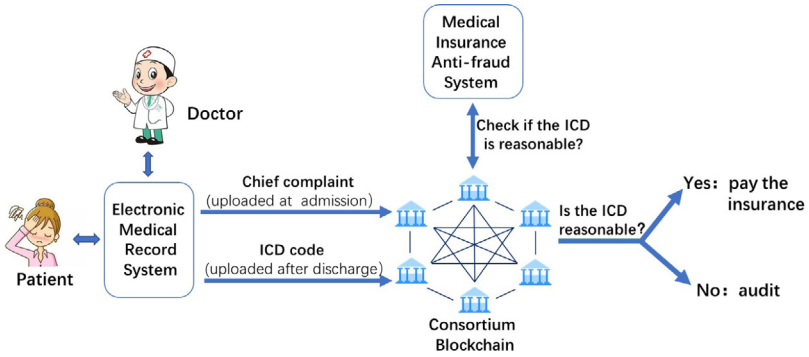


Fig. 1. The proposed anti-fraud framework.

be considered as a fraud and need to be manually audited. The value of k can be determined according to the actual situation, and a higher k means less-strict auditing.

Text classification is an important task in Natural Language Processing (NLP). NLP algorithms based on traditional machine learning mainly rely on hand-crafted features, which are time-consuming and often incomplete. The deep learning-based methods make it possible to automatically learn the feature representation of text. In recent years, the deep learning-based pre-trained language models represented by BERT (Bidirectional Encoder Representations from Transformers) [7] have achieved state-of-the-art in multiple text classification tasks. However, the deep learning-based models are mostly black boxes, which are not explicable. Inspired by study [8,9], we design an explainable BERT-LE algorithm, which combines the pre-trained BERT model with Label Embedding method to classify chief complaints. The proposed method can explain the classification results from Chinese character level to help users better understand the model.

3.1. Inputs and outputs

In most Chinese NLP tasks, word segmentation is required. Considering the lack of a standard Chinese medical glossary, word segmentation is not performed in our task.

In BERT-LE model, the Chinese characters of a chief complaint and all the ICD-10 codes are the model inputs, and the prediction probabilities for all ICD-10 codes are the outputs.

3.2. BERT-LE model structure

Fig. 2 shows the network architecture of the BERT-LE model. There are seven layers in the model, which are input layer, BERT layer, label embedding layer, character local feature extraction layer, character-ICD match layer, fully connected layer, and softmax layer from bottom to top.

3.2.1. Input layer

The input chief complaint must be a fixed-length text sequence. We specify the maximum length n of each input sequence by analyzing the length of the corpus samples. Since the patients' chief complaints are relatively streamlined, n is set to 36 in our study. Sequences shorter than n are padded, and those longer than n are intercepted. Then each Chinese Character x_i is fed to the BERT layer. The input sequence X can be described as Eq. (1).

$$X = \{x_1, x_2, \ldots, x_n\} \tag{1}$$

To jointly train the representations of ICD codes, we also input all ICD codes into the model, and the formula is expressed as follows:

$$Y = \{y_1, y_2, \ldots, y_m\} \tag{2}$$

where m represents the number of ICD codes.

3.2.2. BERT layer

The BERT layer encodes each input Chinese character to an embedded vector using pre-trained BERT model. BERT is a deeply bidirectional, unsupervised language model based on transformer [10]. It is pre-trained using a combination of masked language modeling and next sentence prediction objectives on large corpus to learn contextual representations of words. The formula of BERT layer is expressed as Eq. (3).

$$E_X = BERT(X) \tag{3}$$

Suppose the dimension of the embedded vector is e, the output of the BERT layer is a matrix $E_X \in R^{n\times e}$.

3.2.3. Label embedding layer

The label embedding layer maps the one-hot encoding of each ICD code into an embedding vector. Suppose the embedding size is h, the equation is described as follows:

$$E_Y = W_h \cdot Y = \{e_1, e_2, \ldots, e_m\} \tag{4}$$

Where W_h represents a 2D embedding matrix with dimensions of $m \times h$. The output of embedding layer is a matrix $E_Y \in R^{m\times h}$, and e_i is the embedding vector of the ith ICD code.

3.2.4. Character local feature extraction layer

This layer uses one-dimensional convolution operation to extract local spatial information among consecutive characters. Multiple convolution kernels of different sizes are used in this layer. The size of convolution kernel can be considered as k of k-gram model. Different from Text-CNN [11] model, we use convolution operation to extract local feature representation for each Chinese character, so "same" padding of the input text sequence is applied. Suppose the kernel size is k, the convolution operation can be expressed as formula (5).

$$c_{i1} = w \cdot x_{i:i+k-1} + b_c \tag{5}$$

where c_{i1} is a feature generated from a window of characters $x_{i:i+k-1}$, w is a filter with dimensions of $k \times e$, b_c is bias.

Assuming there are f filters, each Chinese character of the input sequence can generate f features, so the feature of ith Chinese character can be expressed as Eq. (6):

$$c_i = [c_{i1}, c_{i2}, \ldots, c_{if}] \tag{6}$$

Then we concatenate all the features obtained by multiple convolution operations as the final feature representation of each Chinese character. We use matrix $C \in R^{n\times h}$ to represent the output text sequence, where h is the feature dimension of each Chinese character.

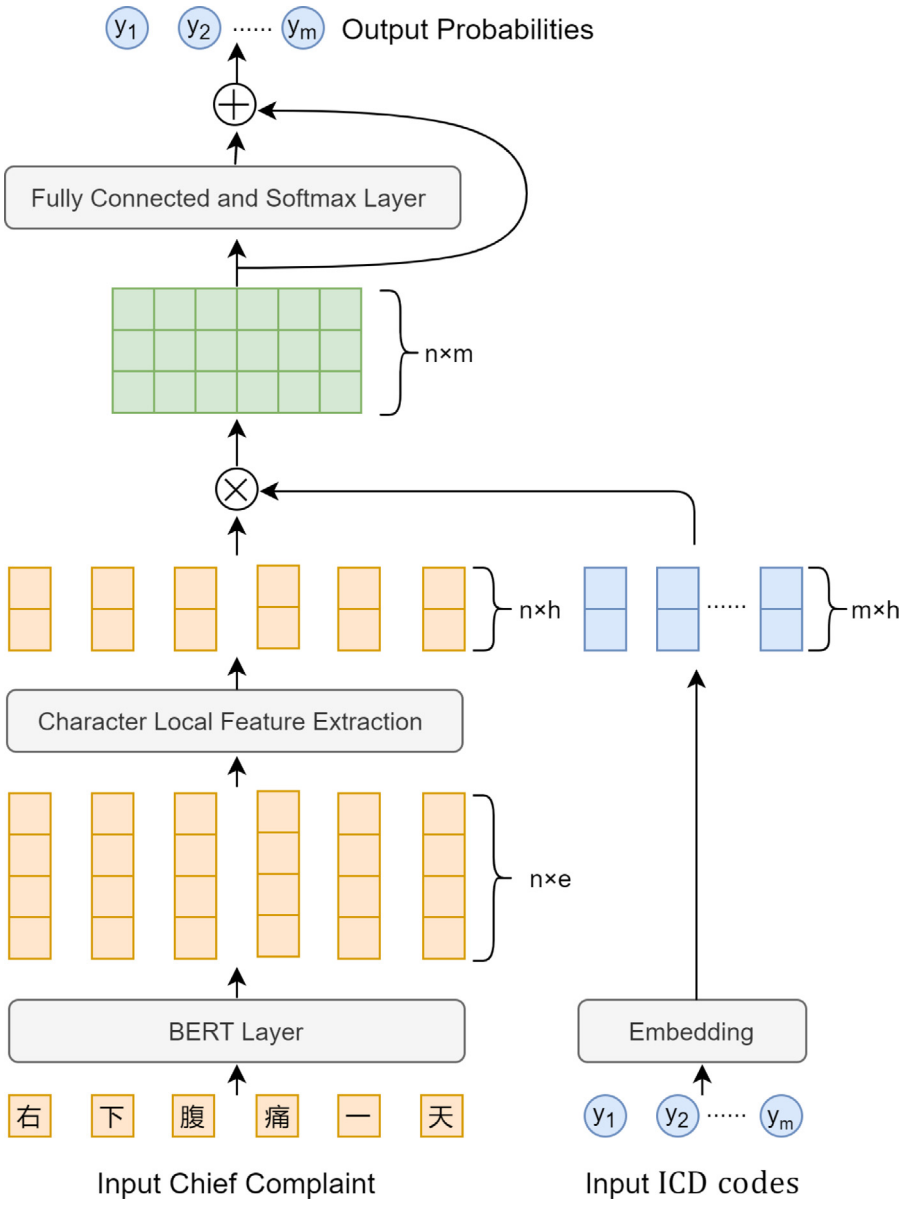


Fig. 2. Architecture of BERT-LE model.

3.2.5. Character-ICD match layer

This layer calculates the matching score between each Chinese character and each ICD code through a dot product operation, the formula is expressed as (7):

$$s_{ij} = c_i \cdot [e_j]^T \tag{7}$$

where c_i is the feature representation of the ith Chinese character in a chief complaint, and e_j is the feature representation of the jth ICD code. In this way, we can obtain the matching score matrix $S \in R^{n \times m}$ of all input Characters and all ICD codes as formula (8):

$$S = C \cdot [E_Y]^T \tag{8}$$

3.2.6. Fully connected layer

The matching score matrix S is passed to a fully connected layer, which is expressed as Eq. (9).

$$F_c = relu(W \cdot M + b_m) \tag{9}$$

Then, a fully connected **Softmax layer** outputs the probability distribution over the ICD-10 categories. We also use residual connections to improve the performance of the proposed deep learning network.

To learn the parameters of the BERT-LE model, the Adam optimizer and cross-entropy loss function are utilized in this paper.

4. Blockchain-based medical records storage and management

Medical records such as chief complaints and ICD codes are important evidence of medical insurance anti-fraud. In order to provide effective protection for medical data anti-tamper, anti-repudiation, security, and integrity, we design a framework of medical data storage and management based on blockchain. In the proposed framework, our major contributions are in the following three aspects:

1. In view of the characteristics of medical data, we design a data storage method combining on-chain and off-chain, which saves storage space and improves the efficiency of blockchain.

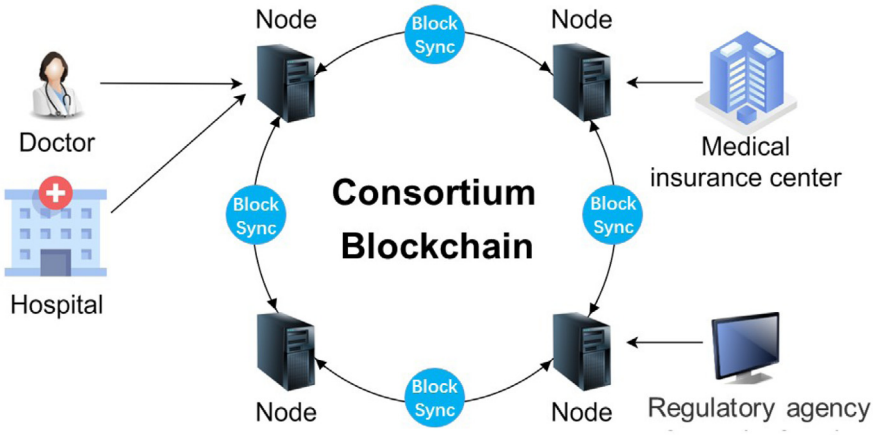


Fig. 3. The framework of medical consortium blockchain.

2. We optimize the PBFT consensus mechanism by dividing blockchain nodes into consensus nodes and application nodes to improve consensus efficiency of blockchain.
3. We design a protocol of medical data storage and query based on smart contracts, which ensure tamper-proofing, traceability, and non-repudiation of medical records.

4.1. Overview of blockchain

Blockchain technology [12] is a mode to manage the generation, access, and use of trusted data through transparency and trust rules. A blockchain is typically managed by a peer-to-peer (P2P) network, the data stored on blockchain are unchangeable, non-forgeable, and traceable. In terms of technical architecture, blockchain is an overall solution composed of distributed storage, block-chain data structure, P2P network, consensus algorithm, cryptography algorithm, game theory, smart contract [13] and other information technologies.

There are three main types of blockchains systems: public, private and consortium blockchains [14]. A public chain refers to a chain in which everyone can participate, even anonymously. A private blockchain is a chain managed and used internally in an organization. A consortium blockchain usually refers to a chain established and governed by multiple parties after reaching a certain agreement or forming a business alliance. Participants need to obtain an invitation or permission to join a consortium blockchain, and all the operations on the blockchain can be controlled with authorities. Considering that the entities of the medical blockchain proposed in this paper are hospitals who are supervised and managed by the government, we adopt a semi-centralized consortium chain to store and manage medical data.

4.2. Data storage and management model

The proposed framework of the medical blockchain is shown in Fig. 3. Users of the blockchain include regulatory agency, medical insurance center, hospitals and doctors. An electronic medical record is written by a doctor. The doctor needs to digitally sign the medical record and initiate a transaction to upload it to the blockchain. Then the doctor's hospital will check the medical record and digitally sign it for confirmation on the blockchain. Medical insurance center queries data from the blockchain to judge the reasonability of disease diagnosis and audits suspicious medical records. Regulatory agency can be connected to the blockchain as a node, or interact with the blockchain through interfaces to synchronize all the data for auditing analysis and track the global business process. If any abnormality is found, an instruction with supervisory authority can be issued to the blockchain to control the business process, nodes, accounts, etc.

4.2.1. Data storage of blockchain

To identify and audit medical insurance fraud, the medical data generated during the hospitalization of each patient, including chief complaint, ICD code, admission record, discharge summaries, progress notes, orders and treatments, front sheet of medical records, etc., need to be stored in a secure and tamper-resistant manner on the blockchain. The chief complaint and ICD code are used to determine the reasonability of a diagnosis, and other data can assist to audit abnormal medical records. In order to save storage space and improve the efficiency of blockchain, we use a combination of on-chain and off-chain for data storage. The data size of the chief complaint and ICD code is very small, and they are the key data for diagnosis reasonability judgment, so we store them on the blockchain. All the other medical records are generated into a PDF file, which is stored in the EMR system of the hospital. The index address and hash value of the corresponding PDF file are uploaded to the blockchain to ensure its integrity and immutability.

In the proposed consortium blockchain, a transaction needs to be initiated by invoking a smart contract for data storage and query. The data structure of a transaction is shown in Fig. 4, which mainly includes transaction number, transaction type, transaction initiating account, transaction receiving account, transaction signature, timestamp, and transaction data. For the uploading transaction of a medical record, the structure of transaction data includes hospital ID, doctor ID, medical record ID, data type and data content. There are three types of data content, namely chief complaint, ICD code, address and hash value of EMR PDF file.

4.2.2. Consensus and node management of blockchain

The Practical Byzantine Fault Tolerance (PBFT) [15] consensus algorithm, which has high efficiency, is adopted in the proposed consortium blockchain. The PBFT uses algorithms based on cryptography such as signatures, signature verification, and hashing to ensure tamper-proof, anti-counterfeiting, and non-repudiation during message transmission. The blockchain can reach a distributed consensus correctly even if there are $(n-1)/3$ (n denotes the number of nodes) malicious or faulty nodes.

The performance of PBFT algorithm will gradually decline as the number of consensus nodes increases, which cannot meet the demand for more nodes to join the chain. In order to improve the efficiency of consensus, we divide the blockchain nodes into consensus nodes and application nodes. The consensus nodes

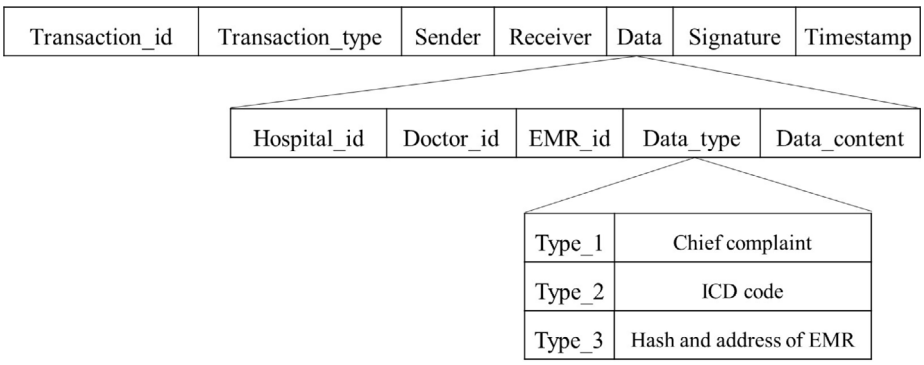


Fig. 4. The data structure of a blockchain transaction.

store all the data of the blockchain and make consensuses for transactions, and they are deployed in the top 20 large hospitals (Large hospitals usually have better network and security guarantees.), regulatory agency, and medical insurance center. The application nodes are deployed in each hospital, they do not participate in consensus, only synchronize and verify the ledger.

4.3. Data storage and management process

The process of medical records storage and management based on blockchain includes four stages: user registration, data storage and query, reasonability verification of diagnosis, and manual auditing.

4.3.1. User registration

In order to trace the behaviors on the chain to specific users, the consortium blockchain uses a permission mechanism to realize the identification and creditability of users, and only users who have been authenticated can join the blockchain. The regulatory agency first run an initialization algorithm to generate the master key and public parameters. Then each user submits registration information (such as doctor's identity information, business certificate of the hospital, etc.) to the regulatory agency for registration. The regulatory agency verifies the identity of each user, and generates a pair of public and private keys (U_PK, U_SK) and a public-key certificate U_Cert for each authenticated user. The string calculated from the user's public key by a one-way hash algorithm is used as the account address of the user, denote as $U\_Address$. The private key that known only by the user corresponds to the password of a traditional authentication system. Finally, the public key U_PK, private key U_SK, digital certificate U_Cert, and account address $U\_Address$ are sent to the user through a secure channel. In the consortium blockchain, a user becomes a legal entity after registration and authentication, and his/her account is identified by the account address. When a user submits a transaction to the consortium chain, he/she needs to sign the transaction with his/her private key, and then the blockchain nodes verify the digital signature according to the user's digital certificate. In this way, the transactions of blockchain can be controlled and traced back to specific users. The registration information of a user can be expressed as follows:

$$User\ \{U\_PK, U\_SK, U\_Cert, U\_Address\} \tag{10}$$

4.3.2. Data storage and query

We use smart contracts to control the storage and query of data on the blockchain, the detailed process is described as follows:

(a) Smart contract deployment. All the data related parties (doctors, hospitals, medical insurance center) agree on the requirements for the data to take effect, and then the regulatory agency creates and deploys smart contracts.
(b) Data upload. The doctor writes a medical record and digitally signs it using his/her U_SK and U_Cert, then the signed record is uploaded to blockchain by invoking a smart contract to initiate a transaction. When the medical record is uploaded successfully, the doctor's hospital will be notified to confirm the medical record.
(c) Data confirmation. The hospital confirms and signs the medical record on the chain.
(d) Data query. The medical insurance center queries the medical record and verifies the validity and integrity of the signature to ensure the authenticity and traceability of data.

The process of storing data on the blockchain is guaranteed by smart contracts. No party can change the data on the blockchain, only data appending is allowed. To avoid impacting the hospital's business, the proposed solution uploads data to the blockchain in a quasi-real-time asynchronous manner. In addition, we use smart contracts to control access to data. The Doctors can only read the data they have signed and uploaded. Hospitals can query the data uploaded by their doctors. The medical insurance center and the regulatory agency have access rights to all data.

4.3.3. Reasonability verification of diagnosis

The diagnosis verification contract needs to be invoked by the hospital (or medical insurance center) to initiate a transaction when the reimbursement settlement is made. The smart contract calls the diagnosis verification service deployed outside the blockchain through Oracle [16] to determine whether the diagnosis is reasonable, and writes the results into the blockchain. The Oracle and reasonability verification service are distributed deployed to ensure the credibility of off-chain data. If the diagnosis is reasonable, the medical insurance center will pay the fee directly; otherwise, a manual review will be carried out.

4.3.4. Manual auditing

During manual auditing, the staff of the medical insurance center queries the relevant medical records and their signatures from the blockchain, and traces them back to related doctors and hospitals. The audit results shall be signed by doctors, hospitals and medical insurance center and stored on the blockchain. The doctors and hospitals who have medical insurance fraud behaviors will be punished, and the frauds will be recorded in their credit files. The medical insurance center could impose further restrictions on doctors and hospitals with poor credit.

Table 1
Statistics of the datasets used in the experiments.

Statistics of the dataset	hospital-A	hospital-B
Number of records in the dataset	198810	434319
Maximum length of chief complaint	38	25
Minimum length of chief complaint	2	2
Average length of chief complaint	12	11
Number of ICD-10 codes	130	209
Vocabulary size by Chinese character	1456	1984
Vocabulary size by Chinese word	6760	6775
Number of records in the training dataset	138821	304013
Number of records in the validation dataset	29996	65158
Number of records in the testing dataset	29993	65148

Table 2
1st, 5th, 10th, 20th, 50th, 80th, 110th and 130th ICD-10 codes in terms of their quantity and frequency of appearance in the hospital-A dataset.

Rank	ICD-10 code and description	Number of patients	Frequency of patients (%)
1	K80: Cholelithiasis	7552	3.80
5	N18: Chronic kidney disease	4231	2.13
10	I25: Chronic ischemic heart disease	3839	1.93
20	M35: Other systemic involvement of connective tissue	2299	1.16
50	G50: Disorders of trigeminal nerve	1330	0.67
80	G51: Facial nerve disorders	904	0.45
110	D12: Benign neoplasm of colon, rectum, anus and anal canal	597	0.30
130	O36: Maternal care for other fetal problems	503	0.25

4.4. Security analysis

4.4.1. Data tamper-proofing

The blockchain is comprised of various blocks that are linked with one another. A block of the blockchain contains a timestamp, the hash of previous block and a Merkle root. The Merkle root is generated by the hashes of all the transactions in the block. If any block changes, the hash values of all subsequent blocks will change accordingly. Unless 2/3 of the consensus nodes in the consortium blockchain are tampered with, the tampered data will be invalid because the nodes cannot reach a consensus. In our framework, the consensus nodes are deployed across multiple organizations, and each organization has well-established network security mechanism. It is very difficult to control 2/3 of the consensus nodes at the same time, which ensures that the data cannot be tampered with.

4.4.2. Traceability and non-repudiation

The medical records stored on the blockchain are digitally signed by doctors and hospitals, and verified by the nodes. Digital signatures guarantee that medical insurance fraud can be traced to specific doctors and hospitals and cannot be denied.

5. Experiments

5.1. Datasets and preprocessing

We tested the proposed approach on two datasets from two large 3 A hospitals (hospital-A and hospital-B) in Jiangsu Province of China. The chief complaints of patients were extracted from admission records, and the ICD-10 codes were extracted from the front sheet of medical records. Compared with the diagnostic codes recorded at the time of admission by doctors, the ICD-10 codes in the front sheet of medical records reveals higher accuracy, because the professional coders have revised them according to the complete hospitalization information of the inpatients. In the experiments, we only selected the first three digits of each ICD-10 code.

There are sparsity and imbalance problems in the datasets, because many of the ICD-10 codes are only allocated to a small number of chief complaints. Therefore, the data preprocessing is necessary. The ICD-10 codes whose number is less than 500 were removed. Irregular complaints and complaints related to non-initial admission (e.g., Z08: Encounter for follow-up examination after completed treatment for malignant neoplasm, Z51: Encounter for other aftercare) were also filtered out. Because these chief complaints do not contain symptom information, they are meaningless in predicting a diagnosis according to the chief complaints. After filtering, there are 198,810 chief complaints and 130 ICD-10 codes in the hospital-A dataset, which covered about 80% of all hospitalized patients. In the hospital-B dataset, there are 434,319 chief complaints and 209 ICD-10 codes, which covered nearly 70% of all hospitalized patients.

Table 1 shows the statistics of the two datasets used in the experiments. In the hospital-A dataset, after deleting non-Chinese characters, such as punctuation marks, the maximum length, minimum length, and average length of a chief complaint is 38, 2, and 12, respectively. The vocabulary size of all Chinese characters in the dataset is 1456. We take the popular toolkit JieBa to segment words with a self-defined dictionary, and the vocabulary size at Chinese word-level is 6760. In the hospital-B dataset, the maximum length, minimum length, and average length of a chief complaint is 25, 2, and 11, respectively. The vocabulary size is 1984 and 6775 at the Chinese character-level and word-level separately. To evaluate the proposed algorithm, each dataset was divided into a training set, a validation set, and a test set with a proportion of 70:15:15. The sizes of corresponding sets are shown in Table 1 .

Table 2 shows the 1st, 5th, 10th, 20th, 50th, 80th, 110th and 130th ICD-10 codes in terms of their quantity and frequency of appearance in the hospital-A dataset. The ICD-10 code with the maximum number of patients is K80, and its number is 7,552, which accounts for 3.8% of all patients. The ICD-10 code with the minimum number of patients is O36, and its number is 506, which accounts for 0.25% of all patients.

Table 3 shows the 1st, 5th, 10th, 20th, 50th, 80th, 110th, 140th, 170th, and 209th ICD-10 codes in terms of their quantity and frequency of appearance in the hospital-B dataset. The ICD-10 code with the maximum number of patients is I63, and its number is 40,107, which accounts for 9.23% of all patients. The ICD-10 code with the minimum number of patients is J33, and its number is 501, which accounts for 0.12% of all patients.

It can be seen from the tables that the ICD codes distribution of hospital-A and hospital-B is quite different.

Table 3
1st, 5th, 10th, 20th, 50th, 80th, 110th, 140th, 170th, and 209th ICD-10 codes in terms of their quantity and frequency of appearance in the hospital-B dataset.

Rank	ICD-10 code and description	Number of patients	Frequency of patients (%)
1	I63: Cerebral infarction	40107	9.23
5	C34: Malignant neoplasm of bronchus and lung	9409	2.17
10	K29: Gastritis and duodenitis	6562	1.51
20	G45: Transient cerebral ischemic attacks and related syndromes	4553	1.05
50	M17: Osteoarthritis of knee	1909	0.44
80	K92: Other diseases of digestive system	1394	0.32
110	E28: Ovarian dysfunction	1054	0.24
140	N80: Endometriosis	848	0.20
170	H40: Glaucoma	649	0.15
209	J33: Nasal polyp	501	0.12

5.2. Baseline methods

We compared the proposed BERT-LE model with several benchmarks, which are listed as follows.

- Text-CNN [11]. A classic text classification algorithm based on CNN (Convolutional Neural Network). Text-CNN learns the feature representations of sentences through multiple one-dimensional convolution kernels, which has a strong ability of shallow text feature extraction. It has been widely used in short text classification because of its good effect and fast training speed. We conduct experiments on Text-CNN for Chinese character-level and word-level, respectively.
- DPCNN [17]. Deep Pyramid CNN (DPCNN) is a word-level text categorization method with deeper CNN layers, which can capture the long-distance dependency of a text and achieve better performance compared with many other approaches.
- Text-RNN [18]. Recurrent Neural Network (RNN) is a classical method of sequence modeling, it is good at modeling text sequences with variable length and capturing inherent features of nature language. Researchers have strong motivation to use RNNs over CNNs in many NLP tasks. In our experiments, the RNN network consists of two layers of Long Short-Term Memory (LTSM) Network. The input characters are first embedded, and the embedding outputs are fed to two LSTM layers. The outputs of LSTM layers are then fed to a fully connected layer. Finally, a SoftMax function is applied to obtain the prediction probabilities of ICD-10 codes.
- HAN [19]. Hierarchical Attention Network (HAN) is a type of hierarchical RNN with attention mechanisms. The model divides the text into a certain number of sentences and then applies encoder and attention to words and sentences, respectively. As the sentences of our datasets are short, they are not divided.
- FastText [20]. It is a word vector modeling and text classification tool proposed by Facebook which provides a simple and efficient way for text categorization and feature representation learning. Its performance is comparable to that of deep learning, but at a much faster speed.
- LEAM [8]. Label-Embedding Attentive Model (LEAM) is a novel attention model for text classification. It jointly trains words and labels to obtain the embedding of them in the same vector space. Compared with other deep learning approaches, it requires fewer parameters and has faster convergence speed.
- BERT [7]. We use the Chinese pre-trained model BERT-wwm [21], an updated version of base BERT, as the baseline in our task. It was proposed by Harbin industrial university and performed slightly better than the base BERT model in Chinese text categorization tasks.

5.3. Implementation details

We trained the model on the training set and tuned the hyper-parameters with the validation set. The model parameters are selected based on the classification accuracy of the validation set. The experiments show that our task is insensitive to hyper-parameters, there is little difference between the experimental results of different parameters. Due to space limitations, we only detailed the parameters of hospital-A dataset.

In the experiments, the Text-CNN, Text-RNN, HAN, FastTex, and LEAM models all adopted ADAM optimization algorithm, and set the same maximum length of the chief complaint and word embedding size, with values of 36 and 64 respectively. The training of the models was stopped when the performance did not improve for 10 epochs. The other hyper-parameters of different approaches are described below.

Text-CNN: The convolutional kernel sizes were 2, 3, 4, and 5, the dropout rate was 0.5, the number of filters was 128, the number of neurons of the fully connected layer was 256, the learning rate was 0.001, the batch size was 64.

DPCNN: The convolutional kernel size was fixed to 3, the number of convolutional layers was 3, and all other parameters were identical to those of Text-CNN.

Text-RNN: The dropout rate was 0.2, the number of hidden layers was 2, the number of neurons in the fully connected layer was 128, the learning rate was 0.001, the batch size was 64.

HAN: The number of neurons in the fully connected layer was 256, and all the other parameters were identical to those of Text-RNN.

FastTex: The number of neurons in the fully connected layer was 256, the learning rate was 0.001, the batch size was 64.

LEAM: The convolution kernel size was 3, the number of neurons of the hidden layer was 300, the dropout rate was 0.5, the penalty of tag was 1.0, the learning rate was 0.001, the batch size was 128.

BERT: The maximum sentence length was the same as the other baselines, which was also 36. The batch size was 16. The number of epochs was tuned from 1 to 5, and the learning rate was selected from 5e5, 4e5, 3e5, 2e5, and 1e5.

BERT-LE: We used two convolutional kernels with size 2 and 3 respectively to learn features of characters, the number of filters was 16, the number of neurons of the fully connected layer was 64, the embedding size of ICD code was 32.

5.4. Results and analysis

To compare the performance of proposed BERT-LE with the baselines, we took the average precision, recall, and F1-score as metrics. As we know, a piece of chief complaint may lead to different kinds of diseases. For example, the chief complaint of "having a fever for 2 days" may indicate an acute upper respiratory infection (J06), or viral pneumonia (J12), or other diseases. Therefore, we evaluated the top-k performance. The predicted

Table 4
Average precision, recall, and F1-score for prediction on Hospital-A dataset. Boldface is the best performance.

Methods	Average performance (Top-1)			Average performance (Top-5)			Average performance (Top-10)		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Text-CNN character	65.36	58.89	58.94	94.03	89.13	90.86	97.61	95.71	96.55
Text-CNN word	64.76	58.48	58.53	93.54	88.56	90.21	97.27	95.22	96.10
DPCNN	65.28	58.77	58.80	94.06	89.07	90.75	97.60	95.59	96.47
Text-RNN	63.27	58.47	57.88	93.46	88.70	90.27	97.28	95.48	96.27
HAN	63.57	58.54	58.35	93.77	88.79	90.52	96.98	96.91	96.88
FastText	58.13	52.73	52.31	92.53	84.46	86.58	96.71	93.02	94.50
LEAM	63.17	56.80	55.80	94.06	88.48	90.20	97.56	95.33	96.30
BERT	65.47	60.91	60.75	**94.22**	90.17	91.61	97.70	**96.15**	96.84
BERT-LE	**65.74**	**61.18**	**61.45**	94.14	**90.26**	**91.69**	**97.77**	**96.15**	**96.88**

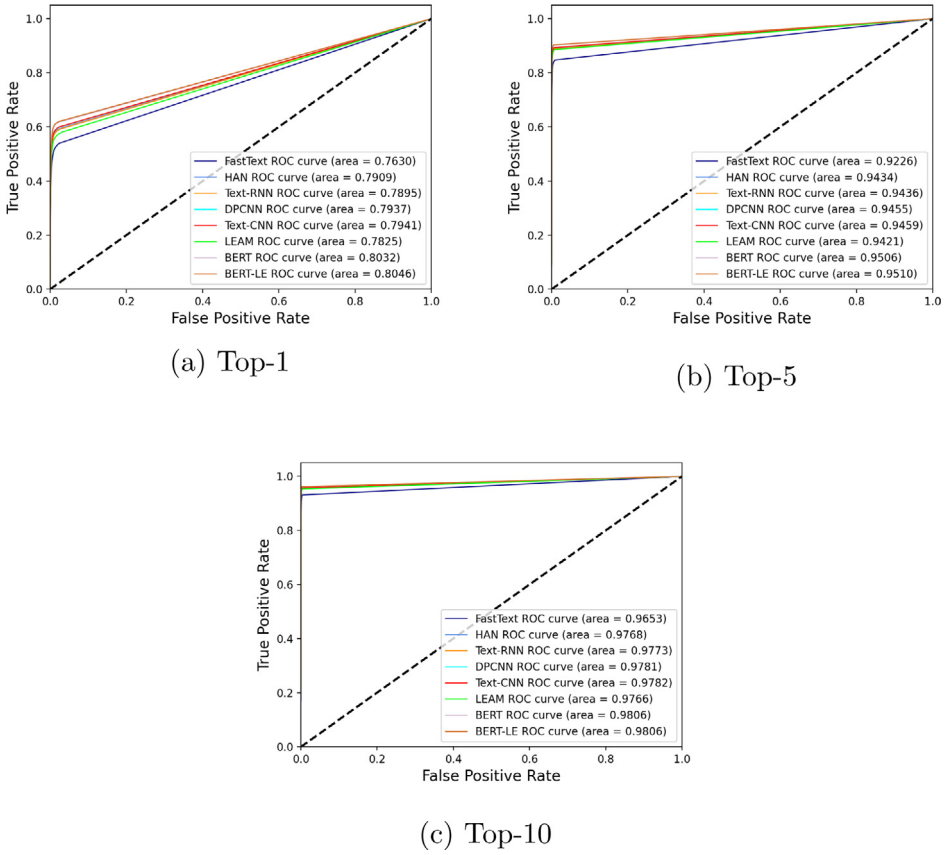


Fig. 5. ROC and AUC of prediction results on Hospital-A dataset.

probabilities of all ICD-10 codes were sorted in descending order. If the truth ICD-10 code of a chief complaint was in the top-k set of the predicted results, the prediction result was considered to be correct. In the experiments, we set the k to 1, 5 and 10. It becomes a traditional text classification problem when k is set to 1.

Table 4 shows the results of all the algorithms on Hospital-A dataset. Fig. 5 shows the ROC and AUC of prediction results. From Table 4 and Fig. 5, we can see that the pre-trained models BERT and BERT-LE are superior to the traditional deep learning approaches; the shallow model FastText performs the worst. The BERT-LE model proposed in this paper achieves the best performance, especially in the top-1 prediction. In the top-5 and top-10 predictions, the performance of BERT-LE does not have many advantages compared with BERT, but BERT-LE has good interpretability, whereas BERT is unexplainable. In the non-pre-trained models, the results of CNN-based methods (Text-CNN, DPCNN) are better than RNN-based methods (Text-RNN, HAN), LEAM is the worst-performing one. Although the network of DPCNN is deeper, its performance has not improved compared with TextCNN, perhaps because the text length of the chief complaint is short, its advantage of mining remote relationship features is not utilized in our task. Compared with traditional RNN, the performance of HAN is slightly improved but the differences are insignificant, which indicates that the attention mechanism has no significant advantage in the short chief complaint text. The performances of word-level Text-CNN is inferior to that of character-level, which may be related to the incorrect recognition of some Chinese medical words.

Table 5 shows the results of all the algorithms on Hospital-B dataset. Fig. 6 shows the ROC and AUC of prediction results. The experimental results are similar to those of hospital-A dataset. The pre-trained models BERT and BERT-LE have similar performance, which are superior to the traditional deep learning approaches, only Text-CNN achieved slightly better results in precision than them. In the non-pre-trained deep learning models, Text-CNN performs the best, LEAM is the worst-performing one. Word-level Text-CNN has slightly better results than character-level in top-1 results, but worse in top-5 and top-10 results.

Table 5
Average precision, recall, and F1-score for prediction on Hospital-B dataset. Boldface is the best performance.

Methods	Average performance (Top-1)			Average performance (Top-5)			Average performance (Top-10)		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Text-CNN character	62.52	50.19	51.63	92.88	83.12	86.67	96.40	90.85	93.28
Text-CNN word	**62.97**	50.52	51.94	92.32	82.60	86.15	96.06	90.38	92.86
DPCNN	60.86	50.07	51.25	92.25	82.94	86.36	96.16	90.61	93.03
Text-RNN	61.26	50.21	50.70	91.73	82.71	85.97	95.78	90.53	92.79
HAN	60.76	50.38	51.47	92.24	82.91	86.38	96.06	90.72	93.06
FastText	53.19	43.49	43.96	91.43	77.13	81.29	95.29	86.44	89.85
LEAM	61.07	47.05	47.50	92.87	81.30	85.12	96.31	89.92	92.59
BERT	62.01	**53.36**	54.51	92.48	**84.72**	**87.74**	96.31	**91.74**	93.77
BERT-LE	62.00	53.24	**54.66**	**92.63**	84.57	87.70	**96.44**	91.69	**93.82**

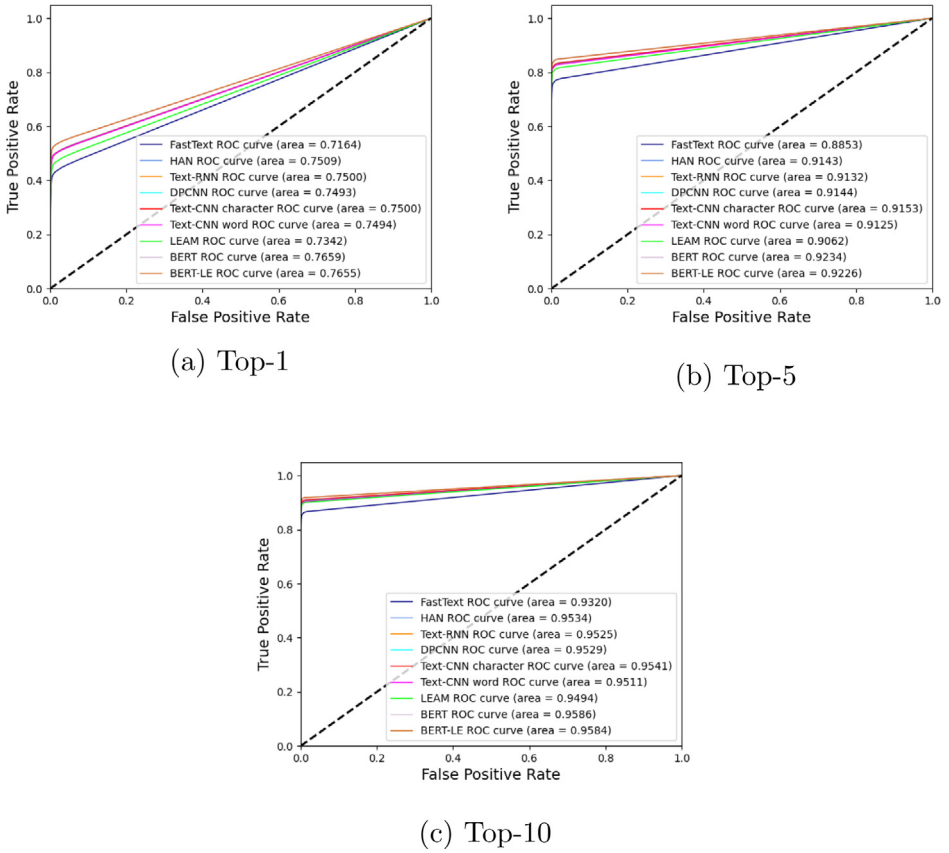


Fig. 6. ROC and AUC of prediction results on Hospital-B dataset.

5.5. Case study

To further evaluate the effectiveness of the proposed framework, we selected 100 records whose predictions were wrong in terms of top-10 from the data and then analyzed them with doctors. We classify the analysis results into three categories as follows:

(1) The chief complaint can barely lead to the ICD-10 code written in the medical record, which indicates that the ICD-10 code is unreasonable and our model is correct.
(2) The chief complaint may lead to the ICD-10 code written in the medical record, but the probability of the ICD-10 code occurrence in the patients is lower than the predicted probability, which indicates that the ICD-10 code may be unreasonable and our model is correct.
(3) The chief complaints may lead to the ICD-10 code written in the medical record, and the probability of the ICD-10 code occurrence in the patients is higher than the predicted probability, which indicates that the ICD-10 code is reasonable and our model is wrong.

The results showed that the first category, the second category, and the third category accounted for 27%, 60% and 13% of the records obtained respectively. The overall effectiveness of the proposed framework is 87% (account by category 1 and category 2). Thus, the proposed framework could play an effective role in anti-fraud of medical insurance and greatly reduce the workload of auditors.

Let us see an abnormal case in the first category. Table 6 describes a chief complaint and the corresponding ICD-10 of a patient. Table 7 shows the top-5 predicted ICD-10 codes with their probabilities. The complaint "Recurrent unconsciousness more than 7 years and aggravated for half a year" has a high probability of leading to brain diseases, such as G40, G45 and I63, which are correctly predicted by our model. However, the actual ICD-10 is D13 (Benign neoplasm of other and ill-defined parts of digestive system), which is obviously unreasonable.

Table 6
The chief complaint and the corresponding ICD-10 of an abnormal case.

Chief complaint	ICD-10 code and description
Recurrent unconsciousness more than 7 years, and aggravated for half a year	D13: Benign neoplasm of other and ill-defined parts of digestive system

Table 7
The top 5 predicted ICD-10s with their probabilities.

ICD-10 code and description	Probability(%)
G40: Epilepsy and recurrent seizures	43.5
R55: Syncope and collapse	14.1
G45: Transient cerebral ischemic attacks and related syndromes	11.9
I63: Cerebral infarction	8.2
I66: Occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction	4.6

左	前	胸	侧	胸	及	背	部	疱	疹	后	疼	痛	一	年	余	ICD code	ICD description
0.16	0.99	0.7	0.4	0.32	1.37	1.64	2.53	5.9	7.6	1.35	0.39	0.31	-1.51	-1.67	-1.32	B02	herpes zoster
-1.17	0.32	0.45	2.35	1.21	0.4	1.15	1.18	-4.06	-5.52	-4.27	-1.48	-1.14	-1.8	-0.83	-0.21	J18	Pneumonia, unspecified organism
-0.01	1.37	0.73	2.66	2.04	1.52	1.06	-3.1	-8.67	-8.35	-2.61	-0.31	-0.09	-0.6	-0.13	0.11	C34	Malignant neoplasm of bronchus and lung
-3.88	-1.32	-1.55	-1.18	-0.36	-1.49	-2.67	-2.47	-1.33	3.63	0.1	-1.77	0.44	0.39	0.05	-0.26	M32	Systemic lupus erythematosus
-1.81	-1.13	0.81	1.85	1.68	1.99	0.07	-2.4	-6.61	-7.24	-4.71	0.15	0.55	-0.25	-0.34	-0.32	I71	Aortic aneurysm and dissection

(a) chief complaint 1

头	痛	四	天	头	晕	伴	恶	心	呕	吐	一	天	ICD code	ICD description
3.19	1.45	0.82	0	2.03	3.43	1.86	0	0	0	1.1	1.35	0.39	I63	Cerebral infarction
3.49	2.72	1.46	0	0.78	2.14	1.27	0	0	0	1.31	1.84	0.41	I61	Nontraumatic intracerebral hemorrhage
2.63	0.94	0	0	1.42	4.52	2.88	0.16	0	0	0.84	1.74	0	G45	Transient cerebral ischemic attacks and related syndromes
3.47	2.69	1.25	0	0.28	0	1.69	0.09	0	0	1.39	1.84	0.49	I60	Nontraumatic subarachnoid hemorrhage
3.27	1.31	0.8	0	0	2.29	0.91	0.46	0	0	0.32	0	0	I67	Other cerebrovascular diseases

(b) chief complaint 2

胸	痛	伴	进	食	梗	噎	感	半	月	余	ICD code	ICD description
0.36	1.14	1.3	0.57	2.39	4.83	3.43	3.47	0.65	0.35	0	C15	Malignant neoplasm of esophagus
0	0	1.52	0.49	1.93	4.19	2.43	3.03	0.68	0.54	0	C16	Malignant neoplasm of stomach
1.52	0.59	0	0.47	1.98	3.54	2.37	1.51	0.04	0.02	0	K22	Other diseases of esophagus
2.89	1.32	1.03	0	0	0	0.32	3.31	0	0.62	0.01	C34	Malignant neoplasm of bronchus and lung
4.01	0.73	0.21	0	0.29	0	0.61	1.22	0.32	0.33	0	I25	Chronic ischemic heart disease

(c) chief complaint 3

Fig. 7. Matching scores of Chinese characters with the top-5 ICD-10 codes.

5.6. Model explainability

To illustrate the explainability of the proposed model, we selected three chief complaints and visualized the matching scores of Chinese characters in the chief complaints with the top-5 ICD-10 codes predicted by our model. It can be inferred that the higher the matching score, the stronger the correlation between the corresponding Chinese character and ICD code. As shown in Fig. 7, the Chinese characters with high matching scores are highlighted in yellow.

In Fig. 7(a), the chief complaint is "左前胸侧胸及背部疱疹后疼痛一年余" (Left thoracic side chest and back pain after herpes for more than one year). The two Chinese characters with the highest matching score with B02 (herpes zoster) are "疱疹" (herpes) in the chief complaint, which is in accordance with the actual situation. "侧胸 (lateral chest) is highly related with J18 (Pneumonia, unspecified organism) and C34 (Malignant neoplasm of bronchus and lung), which is reasonable. Because the symptoms of J18 and C34 are usually reflected in the chest. "疹" (rash) has the highest matching score with M32 (Systemic lupus erythematosus), which is the main symptom of M32. Chest pain is one of the main symptoms of I71 (Aortic aneurysm and dissection), which has the highest matching score with "侧胸" (lateral chest) in the chief complaint.

In Fig. 7(b), the chief complaint is "头痛四天头晕伴恶心呕吐一天" (Headache for 4 days, dizziness with nausea and vomiting for 1 day). Its top-5 predictions are I63 (Cerebral infarction), I61 (Nontraumatic intracerebral hemorrhage), G45 (Transient cerebral ischemic attacks and related syndromes), I60 (Nontraumatic subarachnoid hemorrhage) and I67 (Other cerebrovascular diseases), all of which are brain diseases. In the chief complaint, "头晕痛" (dizziness and headache) and "吐" (vomit) have very high matching scores with these top-5 predictions. What calls for special attention is that the I60 has a high matching score with "头痛" (headache) but a low matching score with "晕" (dizziness). This is also true in practice, as headache is a typical symptom but not dizziness for I60.

In Fig. 7(c), the chief complaint is "胸痛伴进食梗噎感半月余" (Chest pain accompanied by eating obstruction for more than half a month). In the chief complaint, "食" (eating) and "梗噎感" (The food blocked the esophagus and made it difficult to swallow) have very high matching scores with upper gastrointestinal disease C15 (Malignant neoplasm of esophagus), C16 (Malignant neoplasm of stomach) and K22 (Other diseases of esophagus). "胸痛感" (chest pain), which is one of the main symptoms of C34 (Malignant neoplasm of bronchus and lung) and I25 (Chronic ischemic heart disease), has very high matching score with them.

The visualized experimental results reveal that the matching scores between Chinese characters and ICD codes obtained by the proposed model can interpret the classification results very well.

5.7. Limitations

We analyzed several records with incorrect predictions and found that one important reason behind some erroneous predictions was the limitations of the experimental datasets. Since the datasets were obtained from only two hospitals, the proportion

of each ICD-10 code can only reflect the situation of the two hospitals, and the prediction results may be biased. To accurately predict ICD-10 codes and ensure effective implementation of the proposed framework, more data from other hospitals must be collected and trained.

Another limitation of our study is that the proposed model is trained with the chief complaints of inpatients, which cannot reflect the situation of outpatients. Great differences between inpatients and outpatients may be observed in terms of prediction probabilities of diseases.

6. Related work

6.1. Text classification based on deep learning

Deep learning architectures and algorithms have brought about tremendous advances in the fields of computer vision and traditional pattern recognition [22,23]. Following this trend, new deep learning methods are increasingly being used in NLP research. In the past decade, machine learning approaches for solving NLP problems were generally based on shallow models (such as SVM [24]), which were trained on very high-dimensional and one-hot encoding data. In recent years, neural networks based on dense vector representation have yielded good results on a variety of NLP tasks [25]. This trend depends on the success of word embedding and deep learning methods.

Text classification is the basic task of NLP. Due to the great success of deep learning, many advanced algorithms have been proposed in recent years. The text classification methods based on deep learning use word vector to learn semantic representations of words, and then obtain the feature representations of sentences and documents through semantic combination. The semantic combination methods mainly include CNN, RNN, attention mechanism, and GCN (Graph convolutional network), etc. Kalchbrenner et al. [26] proposed a text classification model DCNN based on a two-layers CNN architecture. Zhang et al. [27] proposed a character-level CNN classification method char-CNN. Johnson et al. [17] made references to ResNet [23] and proposed DPCNN, which resolved vanishing gradient problem when increasing network depth. Hong et al. [28] proposed a short text classification model KPCNN by combining knowledge to CNN. Yu et al. [29] proposed a speed-read classification method based on LSTM. Liu et al. [30] proposed the Multi-RNN model based on information sharing mechanism, which adopted the multi-layer LSTM and multi-task training to improve the classification performance of a single task. Lai et al. [31] proposed RCNN to combine CNN and RNN to lean better representations of the words. Xiao et al. [32] proposed char-CRNN, which combined CNN and bidirectional LSTM. The baselines, LEAM [8] and HAN [19], are both attention-based text classification methods. Yao et al. [33] proposed a novel text classification algorithm based on GCN, TextGCN, which models corpus as heterogeneous graph and learns words and document embedding together via GCN. Liu et al. [34] came up with a text classification algorithm based on Bi-GRU and attention mechanism, which learns the content and sequence features of the text, and selects effective features related to labels. In recent years, the pre-trained language models based on transformer have achieved great success in various NLP tasks, and the performance of pre-trained models (such as BERT [7], XLNet [35], RoBERTa [36], GAN-BERT [37], etc.) is significantly better than the non-pretrained deep learning models in text classification tasks. However, the deep learning-based models are mostly inexplicable. The proposed BERT-LE model uses BERT and one-dimensional convolution to simultaneously extract the long-range sequence features and local spatial information of the text, and combines label embedding to learn the feature representations of labels. Our model not only achieves good performance, but also has good explainability.

6.2. Medical data management based on blockchain

As a novel decentralized and trusted solution, blockchain technology has the characteristics of security, reliability, anti-tampering, non-repudiation, and traceability. In the past decade, it has been widely studied and applied in many fields, such as finance, data sharing, supply chain, privacy protection, and healthcare [38–41].

In view of the problems existing in medical data storage and sharing, researchers have carried out a lot of studies based on blockchain technology. Azaria et al. [42] designed a blockchain-based solution, MedRec, to manage and store large amount of medical data in the electronic medical record system. They used Proof of work (POW) as the consensus mechanism, and encouraged data stakeholders (researchers, hospitals, etc.) to participate in the network as "miners" of blockchain. Cao et al. [43] presented a cloud-based eHealth system, which can protect medical records from unauthorized modification and ensure the security, integrity and correctness of data based on blockchain technology. Feng et al. [44] proposed a medical data security model based on the consortium blockchain, in which a hybrid consensus mechanism by combining DPOS (Delegated Proof-of-Stake) and PBFT algorithm was designed. Asma et al. [45] came up with an healthcare management system based on smart contracts of blockchain to manage the storage and access of medical data, which ensured interoperability and auditability of the data handling process. Shen et al. [46] put forward MedChain, a system based on peer-to-peer networks and blockchain to share health data generated from Internet of Things (IoT) devices and mobile applications. Philippe et al. [47] proposed a solution to store and manage medical consent in eHealth systems, the data can be stored in a secure and tamper-resistant manner by using blockchain. Du et al. [48] proposed a business process for medical data anonymous sharing based on blockchain. They designed a novel blockchain consensus mechanism to improve the performance of TPS (Throughput Per Second). Singh et al. [41] put forward a decentralized healthcare management system with blockchain-based electronic healthcare record and smart contracts to ensure that data are secure, reliable and immutable.

To the best of our knowledge, there is currently no research focus on the traceability and non-repudiation of medical insurance fraud based on blockchain.

7. Conclusion

In this paper, we propose a framework to identify fraud of medical insurance based on explainable BERT-LE model and consortium blockchain. It jointly learns the representations of labels and characters to predict the probability of a disease according to a patient's chief complaint, and evaluates the reasonability of the ICD-10 code written in medical record. We also put forward a storage and management process of medical records based on consortium blockchain technology to ensure that data are secure, reliable, immutable, traceable and non-repudiation. The proposed approach can reduce the workload of healthcare insurance auditors and improve the efficiency. The experiments on two real datasets from two 3 A hospitals reveal that our solution can effectively perform anti-fraud for medical insurance and have good explainability.

In future work, we intend to improve the performance of the proposed framework by training more data from other hospitals. Furthermore, utilization of pre-trained word vectors has proven to be effective in our tasks; thus, we aim to train a pre-trained language model on a large medical corpus to improve prediction accuracy of chief complaints. In this study, we predict ICD-10 codes with 3 bits, whereas, a complete ICD-10 code has 6 bits. In

future work, we plan to predict full ICD-10 codes with more medical records to help doctors encode ICD-10 codes more accurately and efficiently. As mentioned in Section 5.7, great differences between inpatients and outpatients may be observed in terms of prediction probabilities of diseases, thus, we also aim to use the chief complains of outpatients to optimize our model.

CRediT authorship contribution statement

Guoming Zhang: Conceived and designed the analysis, Collected the data, Contributed data or analysis tools, Performed the analysis, Wrote the paper. **Xuyun Zhang:** Conceived and designed the analysis, Contributed data or analysis tools, Performed the analysis. **Muhammad Bilal:** Conceived and designed the analysis, Reviewed and edited the paper. **Wanchun Dou:** Conceived and designed the analysis, Reviewed and edited the paper. **Xiaolong Xu:** Conceived and designed the analysis, Contributed data or analysis tools. **Joel J.P.C. Rodrigues:** Contributed data or analysis tools.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

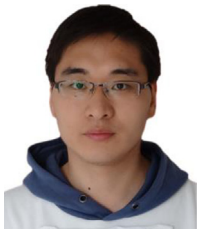
Acknowledgments

This work is supported in part by the National Key R&D Program of China under Grant 2019YFE0190500, the National Natural Science Foundation of China under Grant No. 61672276, Jiangsu Key R&D Program of China under Grant No. BE2019104, the National Key R&D Program of China under Grant No. 2017YFB1400600, and the Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing University. Joel J.P.C. Rodrigues is funded by FCT/MCTES through national funds and when applicable co-funded EU funds under the Project UIDB/50008/2020; and by Brazilian National Council for Scientific and Technological Development - CNPq, via Grant No. 313036/2020-9.

References

[1] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, J. Dean, A guide to deep learning in healthcare, Nature Med. 25 (1) (2019) 24.
[2] I. Mathauer, F. Wittenbecher, Hospital payment systems based on diagnosis-related groups: experiences in low-and middle-income countries, Bull. World Health Organ. 91 (2013) 746–756A.
[3] F. Gao, Y. Tao, Z. Yuan, T. Li, Discussion on the implementation of single disease payment, Chin. Med. Rec. Engl. Ed. 1 (1) (2013) 8–10.
[4] G. Zhang, S. Fu, X. Xu, L. Qi, X. Zhang, W. Dou, An anti-fraud framework for medical insurance based on deep learning, in: International Conference on Advanced Data Mining and Applications, Springer, 2019, pp. 871–878.
[5] M.M. Wagner, W.R. Hogan, W.W. Chapman, P.H. Gesteland, Chief complaints and ICD codes, Handb. Biosurveillance (2006) 333–359.
[6] World Health Organization, International Statistical Classification of Diseases and Related Health Problems 10th Revision, WHO Press, 2011.
[7] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, ACL, 2019, pp. 4171–4186.
[8] G. Wang, C. Li, W. Wang, Y. Zhang, D. Shen, X. Zhang, R. Henao, L. Carin, Joint embedding of words and labels for text classification, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL, 2018, pp. 2321–2331.
[9] C. Du, Z. Chen, F. Feng, L. Zhu, T. Gan, L. Nie, Explicit interaction model towards text classification, in: Proceedings of the AAAI Conference on Artificial Intelligence, 33, 2019, pp. 6359–6366, http://dx.doi.org/10.1609/aaai.v33i01.33016359.
[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, MIT Press, 2017, pp. 5998–6008.
[11] Y. Kim, Convolutional neural networks for sentence classification, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP, ACL, 2014, pp. 1746–1751.
[12] W. Gao, W.G. Hatcher, W. Yu, A survey of blockchain: Techniques, applications, and challenges, in: 2018 27th International Conference on Computer Communication and Networks, ICCCN, 2018, pp. 1–11, http://dx.doi.org/10.1109/ICCCN.2018.8487348.
[13] A. Kosba, A. Miller, E. Shi, Z. Wen, C. Papamanthou, Hawk: The blockchain model of cryptography and privacy-preserving smart contracts, in: 2016 IEEE Symposium on Security and Privacy, SP, IEEE, 2016, pp. 839–858.
[14] B. Zhong, H. Wu, L. Ding, H. Luo, Y. Luo, X. Pan, Hyperledger fabric-based consortium blockchain for construction quality information management, Front. Eng. Manag. 7 (4) (2020) 512–527.
[15] H. Sukhwani, J.M. Martínez, X. Chang, K.S. Trivedi, A. Rindos, Performance modeling of PBFT consensus process for permissioned blockchain network (hyperledger fabric), in: 2017 IEEE 36th Symposium on Reliable Distributed Systems, SRDS, 2017, pp. 253–255, http://dx.doi.org/10.1109/SRDS.2017.36.
[16] R. Mühlberger, S. Bachhofner, E.C. Ferrer, C. Di Ciccio, I. Weber, M. Wöhrer, U. Zdun, Foundational oracle patterns: Connecting blockchain to the off-chain world, in: International Conference on Business Process Management, Springer, 2020, pp. 35–51.
[17] R. Johnson, T. Zhang, Deep pyramid convolutional neural networks for text categorization, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 1, ACL, 2017, pp. 562–570.
[18] A. Graves, Supervised sequence labelling with recurrent neural networks. 2012, 2012, URL http://Books.Google.Com/Books.
[19] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, E. Hovy, Hierarchical attention networks for document classification, in: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016, pp. 1480–1489.
[20] A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, T. Mikolov, Fasttext.zip: Compressing text classification models, 2016, arXiv preprint arXiv:1612.03651.
[21] Y. Cui, W. Che, T. Liu, B. Qin, Z. Yang, S. Wang, G. Hu, Pre-training with whole word masking for Chinese BERT, 2019, arXiv preprint arXiv:1906.08101.
[22] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436.
[23] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE Press, 2016, pp. 770–778.
[24] T. Joachims, Text categorization with support vector machines: Learning with many relevant features, in: European Conference on Machine Learning, Springer, 1998, pp. 137–142.
[25] A. Bakarov, A survey of word embeddings evaluation methods, 2018, arXiv preprint arXiv:1801.09536.
[26] N. Kalchbrenner, E. Grefenstette, P. Blunsom, A convolutional neural network for modelling sentences, 2014, arXiv preprint arXiv:1404.2188.
[27] X. Zhang, J. Zhao, Y. LeCun, Character-level convolutional networks for text classification, in: Proceedings of the 28th International Conference on Neural Information Processing Systems, MIT Press, 2015, pp. 649–657.
[28] S. Hong, J. Oh, H. Lee, B. Han, Learning transferrable knowledge for semantic segmentation with deep convolutional neural network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE Press, 2016, pp. 3204–3212.
[29] K. Yu, Y. Liu, A.G. Schwing, J. Peng, Fast and accurate text classification: Skimming, rereading and early stopping, in: Proceedings of the 6th International Conference on Learning Representations, 2018, pp. 1–5.
[30] P. Liu, X. Qiu, X. Huang, Recurrent neural network for text classification with multi-task learning, in: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, AAAI Press, 2016, pp. 2873–2879.
[31] S. Lai, L. Xu, K. Liu, J. Zhao, Recurrent convolutional neural networks for text classification, in: Proceedings of the 29th AAAI Conference on Artificial Intelligence, AAAI, 2015, pp. 2267–2273.
[32] Y. Xiao, K. Cho, Efficient character-level document classification by combining convolution and recurrent layers, 2016, arXiv preprint arXiv:1602.00367.
[33] L. Yao, C. Mao, Y. Luo, Graph convolutional networks for text classification, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, 2019, pp. 7370–7377.
[34] H. Liu, G. Chen, P. Li, P. Zhao, X. Wu, Multi-label text classification via joint learning from label embedding and label correlation, Neurocomputing 460 (2021) 385–398, http://dx.doi.org/10.1016/j.neucom.2021.07.031.
[35] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R.R. Salakhutdinov, Q.V. Le, Xlnet: Generalized autoregressive pretraining for language understanding, in: Advances in Neural Information Processing Systems, MIT Press, 2019, pp. 5754–5764.

[36] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, 2019, arXiv preprint arXiv:1907.11692.
[37] D. Croce, G. Castellucci, R. Basili, Gan-bert: Generative adversarial learning for robust text classification with a bunch of labeled examples, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 2114–2119.
[38] X. Xu, X. Zhang, H. Gao, Y. Xue, L. Qi, W. Dou, Become: blockchain-enabled computation offloading for IoT in mobile edge computing, IEEE Trans. Ind. Inf. 16 (6) (2019) 4187–4195.
[39] Y.S. Hau, M.C. Chang, A quantitative and qualitative review on the main research streams regarding blockchain technology in healthcare, in: Healthcare, Multidisciplinary Digital Publishing Institute, 2021, p. 247.
[40] X. Xu, D. Zhu, X. Yang, S. Wang, L. Qi, W. Dou, Concurrent practical byzantine fault tolerance for integration of blockchain and supply chain, ACM Trans. Int. Technol. (TOIT) 21 (1) (2021) 1–17, http://dx.doi.org/10.1145/3395331.
[41] A.P. Singh, N.R. Pradhan, A.K. Luhach, S. Agnihotri, N.Z. Jhanjhi, S. Verma, Kavita, U. Ghosh, D.S. Roy, A novel patient-centric architectural framework for blockchain-enabled healthcare applications, IEEE Trans. Ind. Inf. 17 (8) (2021) 5779–5789, http://dx.doi.org/10.1109/TII.2020.3037889.
[42] A. Azaria, A. Ekblaw, T. Vieira, A. Lippman, Medrec: Using blockchain for medical data access and permission management, in: 2016 2nd International Conference on Open and Big Data, OBD, 2016, pp. 25–30, http://dx.doi.org/10.1109/OBD.2016.11.
[43] S. Cao, X. Zhang, R. Xu, Toward secure storage in cloud-based ehealth systems: A blockchain-assisted approach, IEEE Netw. 34 (2) (2020) 64–70, http://dx.doi.org/10.1109/MNET.001.1900173.
[44] T. FENG, Y. JIAO, F. Junli, T. Ye, Medical health data security model based on alliance blockchain, Comput. Sci. 47 (4) (2020) 305–311.
[45] A. Khatoon, A blockchain-based smart contract system for healthcare management, Electronics 9 (1) (2020) http://dx.doi.org/10.3390/electronics9010094, URL https://www.mdpi.com/2079-9292/9/1/94.
[46] B. Shen, J. Guo, Y. Yang, Medchain: Efficient healthcare data sharing via blockchain, Appl. Sci. 9 (6) (2019) http://dx.doi.org/10.3390/app9061207, URL https://www.mdpi.com/2076-3417/9/6/1207.
[47] P. Genestier, S. Zouarhi, P. Limeux, D. Excoffier, A. Prola, S. Sandon, J.-M. Temerson, Blockchain for consent management in the ehealth environment: A nugget for privacy and security challenges, J. Int. Soc. Telemed. EHealth 5 (2017) GKR–e24.
[48] M. Du, Q. Chen, J. Chen, X. Ma, An optimized consortium blockchain for medical information sharing, IEEE Trans. Eng. Manage. (2020) 1–13, http://dx.doi.org/10.1109/TEM.2020.2966832.



Guoming Zhang works at Health Statistics and Information Center of Jiangsu Province. He is now studying for a Ph.D. in Computer Science and Technology at Nanjing University. He has a B.E. in Computer Science and Technology from Shandong University (2006), and an M.E. in Computer Application Technology from Beijing University of Technology (2009). He is a member of Medical and Health Big Data Professional Committee of Jiangsu Province. His research focuses on cloud computing and big data.

Xuyun Zhang (Member, IEEE) received the B.Sc. and M.Eng. degrees in computer science and technology from Nanjing University, Nanjing, China, in 2008 and 2001, respectively, and the Ph.D. degree in computer and information science from the University of Technology Sydney, Ultimo, NSW, Australia, in 2014. He is currently a Senior Lecturer with the Department of Computing, Macquarie University, Sydney, Australia. He also has the working experience with the University of Auckland and NICTA (now Data61, CSIRO). He has so far published authored or coauthored more than 100 refereed academic papers in many high-quality and influential conferences and journals (IEEE Transactions on Parallel and Distributed Systems, IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Computers, IEEE Transactions on Software Engineering, IEEE Transactions on Industrial Informatics (IEEE TII), IEEE Journal on Selected Areas in Communications, and ICDE) in the areas of his research interests, which include scalable and secure machine learning, big data mining and analytics, cloud/edge/service computing and IoT, big data privacy, cybersecurity, etc.

Muhammad Bilal received the B.Sc. degree in computer systems engineering from the University of Engineering and Technology, Peshawar, Pakistan, in 2008, the M.S. degree in computer engineering from the Chosun University, Gwangju, South Korea, in 2012, and the Ph.D. degree in information and communication network engineering from the School of Electronics and Telecommunications Research Institute (ETRI), Korea University of Science and Technology, in 2017. He was a Postdoctoral Research Fellow at Smart Quantum Communication Center, Korea University, Seoul, South Korea, in 2017/2018.

Currently, he is an Assistant Professor with the Division of Computer and Electronic Systems Engineering, Hankuk University of Foreign Studies, Yongin, South Korea. His research interests include design and analysis of network protocols, network architecture, network security, IoT, named data networking, Blockchain, cryptology, and future Internet. Dr. Bilal has served as a reviewer of various international journals, and also served as a Technical Program Committee Member on many international conferences including IEEE VTC, IEEE ICC, Infocom and IEEE CCNC. He is an editor of IEEE Future Directions Ethics and Policy in Technology Newsletter and IEEE Internet Policy Newsletter.

Wanchun Dou (Member, IEEE) received the Ph.D. degree in mechanical and electronic engineering from the Nanjing University of Science and Technology, China, in 2001. He is currently a Lecturer with the Nanjing University of Science and Technology. He is also a Full Professor with the State Key Laboratory for Novel Software Technology, Nanjing University. From April 2005 to June 2005 and from November 2008 to February 2009, he visited the Departments of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong, respectively, as a Visiting Scholar. He has published more than 100 research papers in international journals and international conferences. His research interests include workflow, cloud computing, and service computing.

Xiaolong Xu (Member, IEEE) received the Ph.D. degree from Nanjing University, China, in 2016. He worked as a Research Scholar with Michigan State University, USA, from April 2017 to May 2018. He is currently a Professor with the School of Computer and Software, Nanjing University of Information Science and Technology. He has published more than 80 peer-reviewed papers in the international journals and conferences, including the IEEE TRANSACTIONS ON INDUSTRIALINFORMA TICS, the IEEE TRANSACTIONS ON CLOUD COMPUTING, the IEEE TRANSACTIONS ON BIG DATA, the IEEE INTERNET OF THINGS, the IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTA TIONAL INTELLIGENCE, the IEEE TRANSACTIONS ON COMPUTA TIONAL SOCIAL SYSTEMS, the Journal of Network and Computer Applications, Society of Petroleum Engineers, WWWj, IEEE ICWS, and ICSOC. His research interests include fog computing, edge computing, the Internet of Things, cloud computing, and big data.

Joel J. P. C. Rodrigues [Fellow, IEEE & AAIA] is with Senac Faculty of Ceará, Brazil, head of research, development, and innovation; and senior researcher at the Instituto de Telecomunicações, Portugal. Prof. Rodrigues is an Highly Cited Researcher, the leader of the Next Generation Networks and Applications (NetGNA) research group (CNPq), an IEEE Distinguished Lecturer, Member Representative of the IEEE Communications Society on the IEEE Biometrics Council, and the President of the scientific council at ParkUrbis – Covilhã Science and Technology Park. He was Director for Conference Development - IEEE ComSoc Board of Governors, Technical Activities Committee Chair of the IEEE ComSoc Latin America Region Board, a Past-Chair of the IEEE ComSoc Technical Committee (TC) on eHealth and the TC on Communications Software, a Steering Committee member of the IEEE Life Sciences Technical Community and Publications co-Chair. He is the editor-in-chief of the International Journal of E-Health and Medical Communications and editorial board member of several high-reputed journals (mainly, from IEEE).

He has been general chair and TPC Chair of many international conferences, including IEEE ICC, IEEE GLOBECOM, IEEE HEALTHCOM, and IEEE LatinCom. He has authored or coauthored about 1000 papers in refereed international journals and conferences, 3 books, 2 patents, and 1 ITU-T Recommendation. He had been awarded several Outstanding Leadership and Outstanding Service Awards by IEEE Communications Society and several best papers awards. Prof. Rodrigues is a member of the Internet Society, a senior member ACM, and Fellow of AAIA and IEEE.