**INTRODUCTION**

WITH the growing cases of chronic diseases, more and more patients have to test their health conditions constantly at hospitals, which leads to skyrocketing costs for healthcare systems [1]. To reduce the healthcare costs and improve the healthcare quality, health monitoring systems, which are often built by utilizing decision tree classification, help patients to test their health conditions periodically [2]. Coupled with recent advances of wearable devices and mobile communication networks [3], health monitoring systems work as follows: a hospital first utilizes decision tree classification technique to produce a clinical decision model, and later tests clients’ biomedical data collected from wearable devices and provides decisions for clients based on the model [4], [5]. To further reduce the costs on the hospital side and enable practical deployment, the hospital often outsources the health monitoring services to a cloud server, which brings prominent benefits for both clients and the hospital, such as ubiquitous access, ease of management, and scalability [6].

Despite the well-known benefits, outsourcing health monitoring services to a semi-trusted cloud also arises critical privacy concerns [7]–[9]. On the hospital side, since the hospital may invest a large number of resources to gather sensitive biomedical dataset and train the clinical decision model, the model is valuable intellectual property, which brings commercial benefits to the hospital. Thus, there is a demand for the hospital to protect the content of clinical decision model when outsourcing health monitoring services to the cloud service provider. On the clients’ side, both the physiological features and the clinical decision are sensitive biomedical data, because accidental leakage of either information may reflect the clients’ health condition and lead to serious issues. For instance, if a client has a certain chronic disease, the exposure of health condition deterioration may increase the health insurance costs to the client. With the aforementioned privacy concerns, both the clinical decision model and the biomedical data should be concealed from the cloud service provider in health monitoring systems.

To protect the confidentiality of both clinical decision models and biomedical data, several privacy-preserving decision tree classification schemes have been proposed[10]– [20]. Most of the existing schemes are constructed based on homomorphic encryption (HE) [10]–[14] and secure multiparty computation (MPC) [15]–[19]. HE-based schemes enable privacy-preserving decision tree classification by homomorphically Enrypting the clinical decision model and data, which may incur prohibitive computational overheads [10]– [13]. MPC-based schemes enable multiple parties jointly and privately classify data according to decision trees, but they may lead to expensive communication costs [15]–[19]. To reduce the computation and communication overheads, Liang et al. proposed a secure decision tree classification scheme by utilizing symmetric key encryption [20]. Although the scheme in [20] achieves O(1) computational complexity, it constructs huge indexes, whose size is exponential to the size of decision tree, for privacy-preserving decision tree classification, which incurs heavy storage overheads. In summary, two main challenges should be addressed when designing privacy-preserving decision tree classification schemes for health monitoring systems: (1) Confidentiality: both biomedical data and clinical decision modelshould be protected against the cloud service provider; (2) Efficiency: the computation, communication, and storage costs should be low.

In this paper, we address the aforementioned two challenge simultaneously and propose an efficient and privacy preserving decision tree classification scheme (PPDT) for health monitoring systems. First, we utilize the scheme in [21] to extract rules from decision trees by traversing al decision paths from the root node to the leaf nodes. Then, we build indexes for these rules. The indexes are constructed from Boolean vectors, whose size is polynomial to the number of internal nodes, leaf nodes, and input domains. With such indexes, the decision tree classification process achieves O(1 computational complexity as well as high communication and storage efficiency. After that, we propose PPDT, which incorporates symmetric key encryption, pseudo-random function, and pseudo-random permutation, to enable privacy-preserving decision tree classification by encrypting the aforementioned indexes. Accordingly, PPDT not only protects the confidentiality of both the clinical decision model and biomedical data, but also achieves computation, communication, and storage efficiency for health monitoring systems. The contributions of this paper are summarized as follows.

.We propose an efficient and privacy-preserving decision tree classification scheme (PPDT) for health monitoring systems. First, we transform decision tree classifiers to Boolean vectors, which are indexes that enable O(1) computational complexity for decision tree classification. With such boolean vectors, PPDT significantly improves computation, communication, and storage efficiency simultaneously. By utilizing symmetric key encryption, pseudo-random functions, and pseudo-random permutations to protect the confidentiality of clinical decision models and biomedical data, PPDT significantly reduces the computational costs due to the adoption of low complexity cryptographic primitives.

.We formulate a security definition and give a simulation based security proof for PPDT. First, we identify a leakage function L, which includes the size pattern, search pattern, and access pattern of PPDT. Then, we formulate the L-security definition, which is defined based on the leakage function L. Finally, we provide a simulation-based security proof to demonstrate that PPDT captures the L-security definition. Namely, both the clinical decision model and biomedical data are well protected.

. We conduct performance analyses and evaluations for PPDT. We analyze the computational costs and index sizes of PPDT and the scheme in [20] (SDTC). Despite both PPDT and SDTC are with O(1) computational complexity, the comparison results show that PPDT requires lower computational costs and smaller index sizes than SDTC. The experimental evaluations in Breast-Cancer-Wisconsin dataset also illustrate the performance advantages of PPDT. The performance evaluations demonstrate that: (1) the computational complexity of PPDT is O(1), (2) PPDT only requires micro seconds level execution time, kilobyte-level communication costs, and kilobyte-level storage costs for achieving privacy preserving decision tree classification, and (3) The performance (including computation, communication, and storage efficiency) of PPDT is orders of magnitudes boosted than SDTC.

The remainder of this paper is organized as follows. Section II describes the related work. Section III provides the system model, threat model, and design goals. Section IV illustrates the preliminaries. Section V describes the construction of PPDT. Section VI formulates the leakage function and security definition, and provides a simulation-based security proof. Section VII analyzes and evaluates the performance of PPDT Section VIII concludes this paper.