Decentralized Machine Learning for Intelligent Health Care Systems on the Computing Continuum

Dragi Kimovski
Institute of Information Technology, University of Klagenfurt

Sasko Ristov
Institute of Computer Science, University of Innsbruck

Radu Prodan
Institute of Information Technology, University of Klagenfurt

Abstract—The introduction of electronic personal health records (EHR) enables nationwide information exchange and curation among different health care systems. However, the current EHR systems do not provide transparent means for diagnosis support, medical research, or can utilize the omnipresent data produced by the personal medical devices. Besides, the EHR systems are centrally orchestrated, which could potentially lead to a single point of failure. Therefore, in this article, we explore novel approaches for decentralizing machine learning over distributed ledgers to create intelligent EHR systems that can utilize information from personal medical devices for improved knowledge extraction. Consequently, we proposed and evaluated a conceptual EHR to enable anonymous predictive analysis across multiple medical institutions. The evaluation results indicate that the decentralized EHR can be deployed over the computing continuum with reduced machine learning time of up to 60% and consensus latency of below 8 seconds.

1. Introduction

Many societies and cultures perceive that sexually transmitted diseases (STD) only affect the "others", who follow "sinful" lifestyles and practices. Therefore, discrimination and stigmatization are common outcomes of such distorted depictions of the STD, especially related to the acquired immunodeficiency syndrome (AIDS/HIV). On a global level, the fear of stigmatization prohibits effective disease identification, prevention, care, and treatment adherence, negatively influencing many communities' quality of life [1].

The introduction of the **electronic personal health records** (EHR) systems is a first step towards addressing these issues, especially for illness-related stigmatization. The purpose of the EHR systems is to provide information exchange and curation among different national health care systems and support diagnosis quality, safety monitoring, or medical research.

Although the EHR systems promise substantial benefits for improved care and re-

duced healthcare costs, there are unforeseen difficulties related to privacy and limited diagnosis-related functionalities. The essential barriers that limit the usability and application of the EHR systems are:

- Utilization of central control and orchestration, which could potentially lead to a single point of failure, exposing of private information and hindered interoperability;
- Storing personal data in the hands of a single institution, hindering the data's privacy and limiting knowledge extraction and processing;
- Limited integration with available personal medical **Internet of Things (IoT)** devices, such as heart rate monitors or blood sugar sensors;
- Inability to utilise highly heterogeneous set of computing resources.

Consequently, the EHRs do not allow intelligence to be injected into the process of medical data analysis. This is primarily due to the lack of basic approaches for supporting decentralized management and transparent integration with medical IoT devices.

Recently, the so-called **computing continuum** [2] that federates the cloud services with emerging fog and edge resources, presented a relevant computing alternative for supporting the next generation EHR systems. The computing continuum provides a vast heterogeneity of computational and communication resources, which can allow low-latency communication for fast decision making close to the data sources and substantial computing resources for a complex data analysis. The distributed nature of the computing continuum, further embrace the utilization of **machine learning (ML)** for the creation of intelligent systems and their federation through **distributed ledger technologies (DLT)**. It, therefore, promises transparent integration of the omnipresent IoT data, coming for various personal medical sensors. These technologies promise to be the next disruptive ones and eventually enable intelligently controlled healthcare systems with better societal involvement. The execution of ML over the computing continuum and the integration with

DLTs can make the EHR system personalized, enable transparent integration of IoT devices, and urge active participation of the patients and the medical professionals in the healthcare system. Ultimately, it opens the possibility for training ML algorithms for predictive analysis with medical data belonging to one patient and using the trained models to aid another patient's treatment.

We, therefore, discuss in this article intelligent computational approaches for anonymous analysis of medical information across the computing continuum by creating a decentralized ML overlay for model training in a DLT network.

To support such a decentralized EHR system, we explore the possibility for:

- Creating a decentralized ML network with multi-party computation for secure non-proprietary model training;
- Cross-patient predictive analysis for therapy assessment and research with data acquired from medical IoT devices;
- Transparent orchestration of the EHR system over heterogeneous resources across the computing continuum.

As a proof-of-concept, we propose a decentralized conceptual EHR system that uses ML models for anonymous predictive analysis and evaluate it on a real-world computing continuum infrastructure.

2. Related work

2.1. DLTs and decentralized ML for Healthcare Applications

The support for the future decentralized platforms for medical data analysis with autonomous practices is being researched extensively. In the literature, the work [3] proposes a cross-institutional healthcare predictive model for quality improvement initiatives by predicting the risk of re-admission of a group of patients using data from multiple institutions. This approach sets the ground for developing privacy-preserving ML technology in a DLT. Furthermore, [4] provides an initial medical data management approach through DLT, empowering patients and fighting counterfeit drugs in the pharmaceutical industry.

2

Recently, a feasibility study, presented in [5], explores the idea of applying federated learning for secure multi-institutional data analysis with multiple local models coordinated by a centralized aggregation server. Although the concept is promising, it still requires a centralized model to gather all updates prone to failures and centralised decisions.

OmniPHR [6] proposed a novel DLT-based architecture for distributed and interoperable EHR architecture. The approach allows unified viewpoint of the personal medical information between the patients and the healthcare providers. Furthermore, [7] describes prototype implementation of the OmniPHR architectural model and presents evaluation of the scalability of the approach in terms of integrated production ready databases with information from 40,000 adult patients.

The industry has also explored the utilization of blockchain for private data storage and management. GemOS [8] provides a platform for discovering and sharing disparate data tied to unique identifiers, enabling connection of data sources from different systems on a common ledger and creating proofs-of-existence with verifiable data integrity.

2.2. Decentralized ML across the computing continuum

Limitations in the hospital infrastructures' computational capabilities pose serious problems for deploying ML systems in a decentralized manner. Therefore, the computing continuum has been recently considered as suitable computing infrastructure, capable to meet the conflicting requirements of the EHR systems.

More concretely, the authors in [9] introduce a novel gradient-based training concept for distributed ML models without external computing services over multiple edge resources. The edge devices train the local models with local data coming from multiple IoT and medical devices, which is finally aggregated on one device. The work in [10] presents a novel tree-based ML algorithm, called Bonsai, for efficient prediction on edge and fog devices, close to the IoT devices, which maintains acceptable prediction accu-

racy while minimizing the model size and the prediction costs. Furthermore, [11] presents a distributed learning approach that complements the cloud for providing privacy-aware and efficient analytics. The algorithm divides the deep learning model into multiple smaller ones, which can be placed on the available edge devices, while maintaining a central part of the model in the cloud. Moreover, [12] presents an application of a deep learning algorithm for medical images analysis that uses fixed-point arithmetic, which can fine-tune the analytic algorithm based on medical image segment and the available computing resources on the device.

3. Technological gaps

Based on the related work analysis we identified three research technological gaps.

3.1. Gap 1: Centralized control of the medical data and the ML models.

Due to privacy concerns, the medical institutions manage their data locally, which often leads to inefficient data propagation. This hinders the possibility of training ML algorithms for predictive analysis with medical data belonging to one patient and using the trained models to aid another patient's treatment. The current attempts to integrate ML and IoT with EHR systems, such as federated learning, which enables training across multiple geographically devices, already yielded promising results [13]. The role of federated learning for EHR systems is twofold: (i) it allows distribution of the training over the computing continuum resources, and (ii) it brings privacy in combination with the multi-party computing approach. However, even though these algorithms are distributed, they are centrally controlled and require maintenance of a centralized model, periodically updated by multiple local algorithms and potentially exposing private information. In such an environment, malicious federated learning actors can compromise the ML model by mimicking a local or contributing learner/model's role. Even worse, the central actor gathering the model updates (such as those used by Google or Amazon) may steer it towards own

personal interests that may be different than those of the contributors, i.e., put biases to the model.

To overcome the identified issues, it is essential to research permissioned DLT protocols for federating a set of ML models, with no need for centralized training and later inference, with three key benefits. Firstly, to improve users' control of the models with a secure relation to their data (or training information). Secondly, to enable control of information ownership, shared further down the federation of ML models through an adaptive state transition modeling. Thirdly, to enable anonymous sharing of parts of the ML models (set of rules or wights) between ML systems, with similar characteristics, in the overlay for further improvements.

3.2. Gap 2: Constrained predictive data analysis with limited IoT integration.

The training of decentralized ML models for predictive analysis of anonymous medical information across DLTs also faces serious challenges. One essential issue is to transparently classify one anonymous patient's medical data among various others through a decentralized collaboration of a set of local ML models (with similar characteristics/algorithms), logically located at different medical institutions. Besides, it is difficult to correlate and analyze millions of anonymous non-contextualized medical records produced by various devices, distributed into different locations with different attributes. In this scenario, it is difficult to identify if the data comes from different patients (or even different sensors belonging to the same patient), affecting the predictive analysis. Furthermore, the feasibility of training decentralized ML models for medical information analysis, research, and its integration with IoT devices has never been explored.

Therefore, it is important to explore decentralized approaches for the federation of ML training with guided analytics. The approach should address the problem of non-contextualized training data aggregation, knowledge extraction, and cognitive learning about users' medical and personal data

in an anonymous manner. This could occur through a seamless coupling of ML predictive analysis algorithms on non-contextualized and anonymous medical information.

3.3. Gap 3: Insufficient computing resources and computationally inefficient DLT and ML solutions for Edge training.

The DLT and ML approaches are known to be computationally demanding [14]. However, in large-scale heterogeneous and fragmented environments where patient data spans across geographical boundaries, the important limiting factors are the insufficient computational resource and technical expertise. Concretely, the hospitals do not own a vast infrastructure, and the utilization of high-performance computing services is not always feasible. Furthermore, the employment of the local hospital infrastructure for decentralized ML training can lead to reduced accuracy of the ML model and errors during predictive analysis, especially if the medical data for training is generated by the IoT devices.

Therefore, we should address scalable approaches for efficient model updates in an ML overlay with an increasing number of learners/algorithms distributed across various physical locations. In practice, we should approach scalability, concerning the available resources across the computing continuum, from various angles such as latency for consensus and transaction validation time (e.g., model update). It is therefore essential to explore if we can sacrifice the ML model accuracy to allow execution on computing continuum resources connected directly to the personal medical IoT devices (such as heart rate or blood saturation monitors) or other medical equipment, which might not be directly accessible over the network.

4. Decentralized EHR system architecture

We propose a conceptual EHR system architecture, named STIGMA (see Figure 1). With the STIGMA system, the medical data always stays at the medical institutions and forms the local ML models, but only after

4

performing anonymization. The medical professionals interact with the system through multi-modal diagnosis equipment, enriched with sensor data from personal IoT devices. The medical institutions can register in the STIGMA EHR systems by utilizing strictly defined protocols for interoperability as defined in [6]. The STIGMA EHR system performs in the following manner:

1) *Data Analysis* instance receives a direct multi-modal data stream (MRI, CT scans, IoT heart rate sensors, EEG sensors ...) of the medical procedures.
2) The data stream is afterward analyzed on the available computing continuum resources.
3) *Data Analysis* filters and anonymizes the data stream, which is then sent to the *Model Training* instance.
4) The *Model Training* instance applies ML algorithms to train a model on the available computing continuum resources.
5) After the model is trained, the *Model Training* instance utilizes the *Distributed Ledger* to registers the model (only as a pointer, without exposing the data) and checks for other suitable registered models.
6) Thereafter, if suitable models are found in the *Distributed Ledger* registry, *Model Training* contacts the model owners directly, i.e., other medical institutions, to receive rolling updates or exchange (share) relevant data for model improvement.
7) The STIGMA EHR system can only perform the rolling updates and the data sharing after a consensus (by voting) is reached among all medical institutions, federated by the *Distributed Ledger*.
8) The information is then used for improving the model, which is used to provide real-time support in diagnosing and therapy assessment and is again registered in the *Distributed Ledger*.

All of the above-mentioned steps are continuously managed and synchronized in a decentralized manner by the STIGMA EHR network. It logically forms a peer-to-peer group

that maintains records on all transactions (model updates, inference performance data, and accuracy). The STIGMA EHR network also contains information for the available computing continuum resources (in terms of computing power and available ML models) at each medical institution. The EHR network, therefore, allows all parties of the system to confirm or reject any piece of data added to it, while no data can be deleted from it. This provides a full history of all transactions appeared on the DLT, giving to the EHR a method to insure the correctness of retrieved information.

4.1. ML overlays with decentralized medical data control

To support the creation of the decentralized STIGMA EHR, depicted in Figure 1, we research a DLT-based overlay for the federation of multiple medical institutions through the following actions, directly related to identified technological gaps:

4.1.1. DLT for a decentralized federation of ML models in an overlay. The STIGMA EHR system uses a permissioned [15] protocol to create an appropriate configurable and modular federating architecture addressing EHR systems' requirements for anonymous ML model updates with full control of the private data that does not leave the hospital infrastructure. It relies on scalable DLT management approaches capable of reaching consensus with a minimal number of communication steps with a limited number of ledgers in a permissioned environment.

4.1.2. Model provenance for decentralized ML. Another important aspect of the STIGMA EHR is data provenance, a key concept for supporting the ML-based analysis over decentralized networks, especially when data from IoT devices are used. It enables efficient access approaches that allow all participating ML models in the ML overlay to maintain a copy of the DLT and ensure the availability of the same version of truth. The DLT only contains the transaction logs referring to the ML model updates' fingerprints,
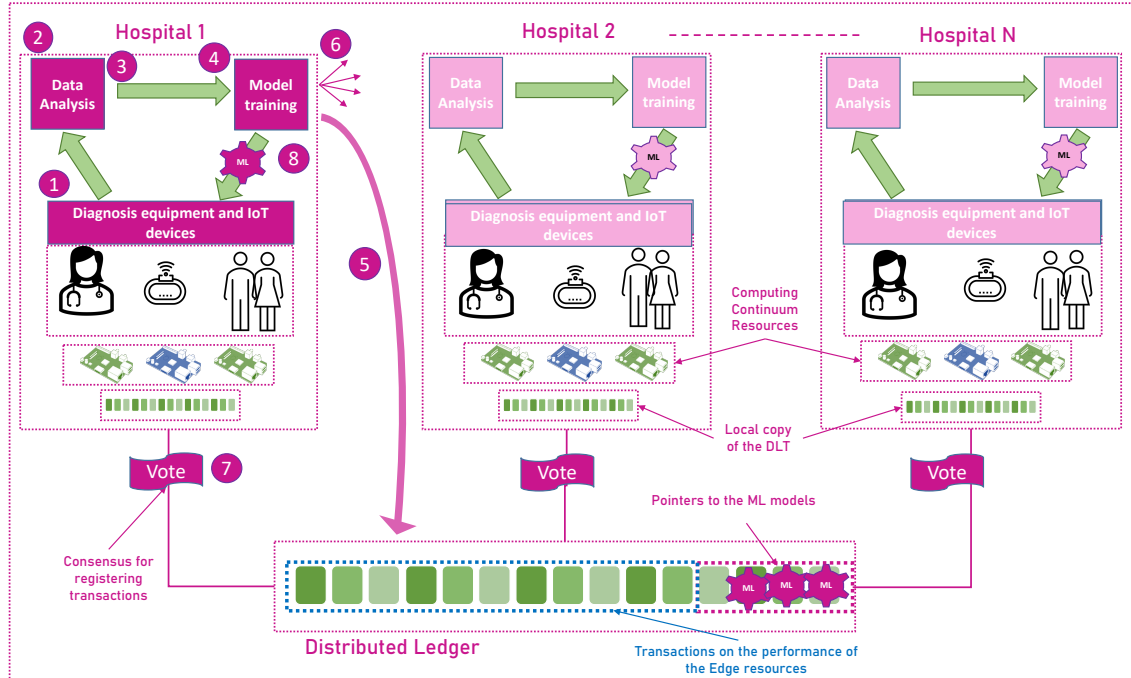
Figure 1: Conceptual architecture of the STIGMA EHR system.

exclusively stored in the hospital computing infrastructures, to reduce replication and network throughput.

4.1.3. Data immutability and secure propagation of decentralized ML model updates with multi-party computation.
This aspect addresses the immutability and propagation of the ML models without violating data privacy during updates. It is used to publish, update and activate anonymous information exchange among the ML algorithms during model training across the overlay. Unfortunately, the current technologies are computationally inefficient, thus not allowing straightforward utilization of the DLTs for complex data sets. We, therefore, utilize the concept of multi-party computation [16] to enable computation on data from different providers. The other participating actors gain no additional information about each other's inputs, except what they learn from the ML model's collaborative output, i.e., decoupling the model from the training data.

4.2. Predictive data analysis with IoT integration.

Another important enabler for the deployment of the STIGMA EHR system is the cross-medical data analysis for improved diagnosis and therapy assessment through distributed ML with IoT medical devices integration. Therefore, we rely on the following solutions:

4.2.1. Predictive analysis with decentralized ML.
The STIGMA EHR system utilizes approaches for automated ML reasoning with a distributed non- and cross-referenced data received from professional medical equipment and personal medical IoT devices. This process reduces uncertainty and mistrust by double-checking the validity of the information and their sources in potentially unpredictable environments. The approach enables shared knowledge and improves data acquisition from IoT devices.

6

4.3. Scalable ML and consensus on the computing continuum.

Multiple research works, such as Paxos and RAFT [17], agree on a single majority value (i.e., state transaction) with a reduced overhead and power requirements. Unfortunately, they still require large computational resources that a resource-constrained hospital infrastructure cannot provide, thus making the deployment of the STIGMA EHR system challenging. Therefore, we modify the current approaches to make them suitable for execution on computing continuum devices. To achieve this, the STIGMA EHR system assesses the complexity of the ML algorithms and the training data structure to select suitable resources in the computing continuum with higher computational capabilities, close to where the data resides in terms of the network distance. Then, based on the available hospital computational infrastructure, a decision is taken where to conduct the training and identify the accuracy level.

5. Real testbed evaluation

To validate the proposed conceptual EHR system, we deployed DLT-based ML systems on a real-world experimental testbed. We emulate the computing infrastructure of the medical institutions by using adequate cloud, fog, and edge resources, as described in Section 5.1. For the evaluation, we implemented the PAXOS 3-phase commit protocol, where each institution in the DLT network keeps track of the current changes. In order to allow execution on multiple heterogeneous systems, we developed the PAXOS protocol in Java 11.0.

5.1. Physical testbed

We utilized a real computing continuum testbed, located at the University of Klagenfurt named *Carinthian Computing Continuum* (C^3) [18] to emulate a network of multiple medical institutions with limited computing capacities.

The C^3 encompasses heterogeneous resources, provided as containers or virtual machines, in multiple performance categories. We have, therefore, identified a subset of resources, usually available in hospitals (such as fog and private cloud infrastructures), and user specific devices (such as edge clusters composed of low-powered portable devices) to conduct the conceptual evaluation (See Table 1).

Centralized Computing Infrastructures (CCI) consists of virtualized instances provisioned on-demand from the Amazon Web Services (AWS). For the evaluation purposes we selected the `m5a.xlarge` and `c5.large` as general purpose instance powered by AMD EPYC 7000 processors at 2.5 GHz and Intel Xeon Platinum 8000 series processor at 3.6 GHz respectively.

Fog Cluster (FC) comprises resources from the local Exoscale (ES) cloud provider, which enables communication latency of $\leq$ 12 ms and maximal bandwidth of $\leq$ 10 Gbps. For the evaluation purposes we identified `ES Medium` and `ES Large` instances, as described in Table 1.

Edge Cluster (EC) comprises five NVIDIA Jetson Nano (NJN) and 32 Raspberry Pi-4 single-board computers (RPi4). We installed `Raspberry Pi OS` (version 2020-05-27) on the RPi devices. We used `Linux for Tegra (L4T)` for the NJN resources. We utilized a managed 48-port layer-3 HP Aruba switch to interconnect all resources in the Edge cluster. The switch supports 1 Gbps per port with a latency of 3.8 μs and an aggregate data transfer rate of 104 Gbps. The Edge cluster is managed by the Edge Gateway System (EGS) based on a twelve-core AMD Ryzen Threadripper 2920X processor at 3.5 GHz and 32 GB of RAM, which is easily available in many medical and business environments. In cases when there are not sufficient resources available at the EC, the EGS is responsible to partially offload the execution of the compute processes to other computing continuum resources, including FC or CCI.

5.2. Experimental design

We designed four sets of experiments according to the characteristics of the conceptual decentralized EHR system and averaged the results over ten runs for statistical signif-

Table 1: The C^3 testbed configuration.

| Instance / Device	CCI (AWS)		FC		EC		
	`m5a.xlarge`	`c5.large`	Exoscale `Large`	Exoscale `Medium`	EGS	NJN	RPi4
CPU type	AMD EPYC 7000	Intel Xeon Platinum 8180	Intel Xeon Platinum 8180	Intel Xeon Platinum 8180	AMD Ryzen 2920	Tegra X1 and ARM Cortex A57	ARM Cortex 72
CPU clock [GHz]	2.5	3.6	3.6	3.6	3.5	1.43	1.5
Memory [GB]	32	8	8	4	32	4	4
Storage [GB]	120	120	120	120	1,000	64	64
BW [Mb/s]	27	26	65	65	813	450	800

icance:

- **DLT network initialization time** evaluates the initialization time of the EHR network encompassing multiple medical institutions in the range of $\{3, 5, 7, 10\}$. The medical institution that initializes the EHR network is considered as the first leader, where the leader interval is 30 ms and the delay between voting rounds is 100 ms. The medical institutions join the EHR network in regular intervals of 10 s.

- **Consensus time** evaluates the time needed for the network encompassing all medical institutions in the range of $\{3, 5, 7, 10\}$ to reach a consensus on a single value. Similar to the previous experiment, the leader interval is 30 ms and the delay between voting rounds is 100 ms. The consensus time is only measured after the network is fully initialized with all participating institutions.

- **ML training time** evaluates the training process of a *convolutional neural network* for object detection with medical multimodal data from laparoscopic procedures [19] limited to 500 samples. The convolutional network has three layers with a kernel size in the range $\{32, 64, 128\}$ and accuracy of 97%. The ML training time also included the overhead required for transferring the trained model to the device where inference will be performed.

- **Edge accuracy** evaluates the trade-off between the accuracy and the training time for the above described *convolutional neural network* on the computing continuum devices. This experiment compares the execution time for training the neural network with an average accuracy of 85% and 70%.

- **Data transfer time** measures the time for transfer of one MB of raw data between an IoT device, connected to the C^3 infrastructure, and the corresponding destination resource. The transfer time has been measured using the Prometheus monitoring system.

5.3. Results
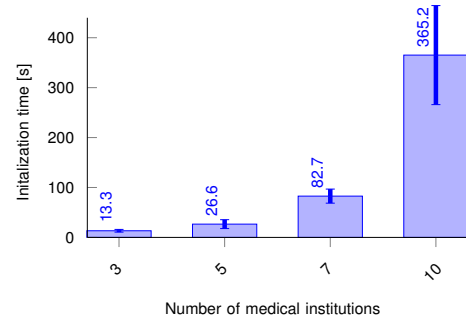
(a) DLT Network initialisation.
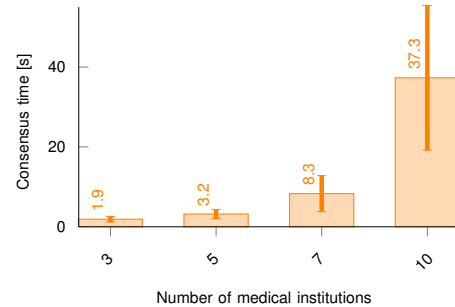
(b) DLT Consensus time.

Figure 2: Consensus evaluation of the STIGMA EHR system.

Figure 2a shows that the current consensus algorithms have limited scalability considering the process of the network initialization. We can observe that the initialization of the EHR network with ten medical institutions can take up to 28 times more time compared to a small network of three institutions, which limits the number of participating institutions in a single decentralized EHR system. However,

8
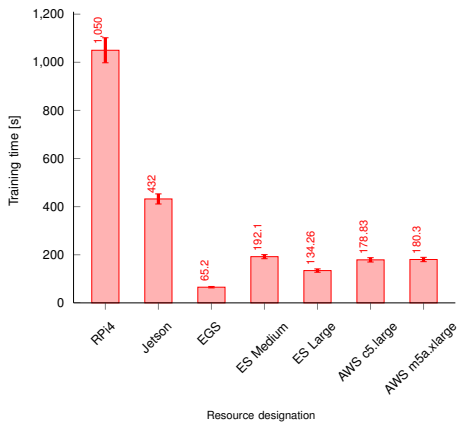
(a) Average training time for accuracy of 97%.
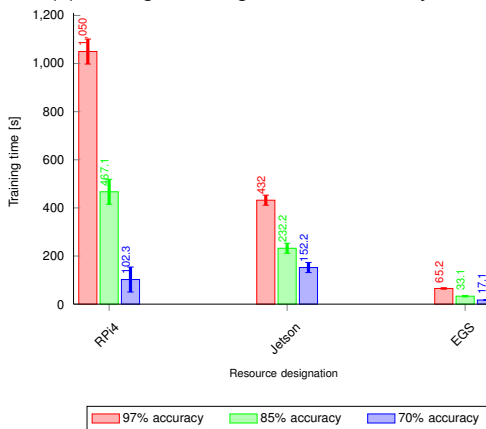
(b) Average training time for lower model accuracy.

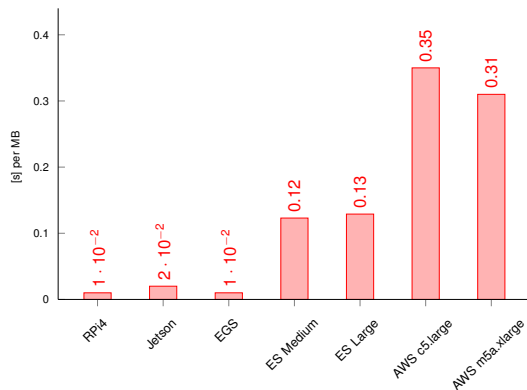Figure 3: ML training in the STIGMA EHR system.

Figure 4: Effective time for transfer of 1 MB of data.

the standard deviation ranges from 29% for ten participating institutions to 58% for three. The reason for the scalability limitation is that all consensus messages must be relayed

through a single coordinator, which although is not a single point of failure, it is a potential performance bottleneck. This is evident during the network initialisation for large number of institutions. However, this experiment proves that up to ten medical institutions can be federated in a single overlay with minimal initialisation overhead.

Furthermore, in Figure 2b we observe a similar trend related to the time needed to reach consensus. The EHR network composed of ten institutions required almost 19 times more time to reach a consensus compared to the small three institutions network. However, we observe a much lower standard deviation, which ranges from 18% for seven participating institutions to 31% for three. Furthermore, compared to proof-of-work approach, implemented in the blockchain protocol, our approach is more efficient in terms or computing resources.

Figure 3a evaluates the suitability of the most commonly available resources for performing ML training over multi-modal medical data. We can observe that the specialized devices for ML, such as the NJN device, are very suitable for performing these tasks and can be easily afforded by medical institutions. Besides, the available EC devices, extended with other resources from the computing continuum, can achieve very low model training times, making them suitable for supporting decentralized EHR systems, especially in cases when the system utilization is low. The reason for this is that the resources across the computing continuum can meet the conflicting requirements of the EHR systems (such as close proximity to the data source and high performance analysis) due to their high heterogeneity.

Figure 3b evaluates the relationship between the accuracy of the ML model and the execution time on the computing continuum. We can observe that reducing the accuracy from 97% to 85% can reduce the execution time by more than 60%. Furthermore, reducing the accuracy to 70% can reduce the execution time on the constrained devices by 90%. However, the trade-off between the accuracy and the execution time depends on

the requirements of the EHR system and the specific medical procedure. In general, this allows for various proximity computing techniques to be applied to improve the performance of the ML training without any significant accuracy penalty.

Lastly, Figure 4 analyses the network performance for raw medical data exchange between the different resources available across the computing continuum. We can observe that the RPi4 and EGS devices can achieve very low data transfer times compared to the CCI and FC instances, which could significantly reduce any computing performance advantage the CCI alone can provide.

6. Conclusion and research directions

To release the potential of the decentralized EHR systems and their transparent support by IoT devices, in this article, we explore the need for the creation of an ecosystem to support the complete lifecycle of medical data sharing and processing. The presented approach enables knowledge extraction for improved medical diagnosis, therapy, and stigma reduction on top of decentralized heterogeneous infrastructures as part of the computing continuum and the IoT environments. We, therefore, identify critical research gaps. Based on the identified considerations, we define concrete research and technical actions required for their implementation. Lastly, we implement STIGMA, a conceptual decentralized EHR system as a proof-of-concept. The system yielded promising results in terms of scalability, which indicate that up to seven different medical institutions can be integrated in decentralized overlay with consensus latency of 8 seconds or lower. In terms of ML learning time we observed that the Edge devices can perform similar to the Cloud resources, and some of them, such as the EGS, can even reduce the training time by 60% compared to the cloud.

Lastly, based on the evaluation results of the conceptual STIGMA EHR system we can conclude that decentralized ML over the computing continuum for medical data analysis can be achieved through the utilisation of scalable consensus algorithms over a per-missioned DLT network with transparent integration of personal IoT devices.

In future, we plan to explore further how we can identify the optimal trade-off between the training accuracy and execution time on low performance devices across the computing continuum.

▪ REFERENCES

1. Sarah E Stutterheim, Lenneke Sicking, Ronald Brands, Ineke Baas, Hilde Roberts, Wim H Van Brakel, Lilian Lechner, Gerjo Kok, and Arjan ER Bos. Patient and provider perspectives on hiv and hiv-related stigma in dutch health care settings. *AIDS patient care and STDs*, 28(12):652–665, 2014.

2. Pete Beckman, Jack Dongarra, Nicola Ferrier, Geoffrey Fox, Terry Moore, Dan Reed, and Micah Beck. Harnessing the computing continuum for programming our world. *Fog Computing: Theory and Practice*, pages 215–230, 2020.

3. Tsung-Ting Kuo and Lucila Ohno-Machado. Modelchain: Decentralized privacy-preserving healthcare predictive modeling framework on private blockchain networks. *arXiv preprint arXiv:1802.01746*, 2018.

4. Matthias Mettler. Blockchain technology in healthcare: The revolution starts here. In *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pages 1–3. IEEE, 2016.

5. Micah J Sheller, G Anthony Reina, Brandon Edwards, Jason Martin, and Spyridon Bakas. Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation. In *International MICCAI Brainlesion Workshop*, pages 92–104. Springer, 2018.

6. Alex Roehrs, Cristiano André Da Costa, and Rodrigo da Rosa Righi. Omniphr: A distributed architecture model to integrate personal health records. *Journal of biomedical informatics*, 71:70–81, 2017.

7. Alex Roehrs, Cristiano André da Costa, Rodrigo da Rosa Righi, Valter Ferreira da Silva, José Roberto Goldim, and Douglas C Schmidt. Analyzing the performance of a blockchain-based personal health record implementation. *Journal of biomedical informatics*, 92:103140, 2019.

8. The GemOS System. https://enterprise.gem.co/gemos/, 2021. [Online; accessed 05-May-2021].

9. Shiqiang Wang, Tiffany Tuor, Theodoros Salonidis, Kin K Leung, Christian Makaya, Ting He, and Kevin Chan. When edge meets learning: Adaptive control for resource-constrained distributed machine learning. In

IEEE INFOCOM 2018-IEEE Conference on Computer Communications, pages 63–71. IEEE, 2018.

10. Ashish Kumar, Saurabh Goyal, and Manik Varma. Resource-efficient machine learning in 2 kb ram for the internet of things. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1935–1944. JMLR. org, 2017.

11. Seyed Ali Osia, Ali Shahin Shamsabadi, Ali Taheri, Hamid R Rabiee, and Hamed Haddadi. Private and scalable personal data analytics using hybrid edge-to-cloud deep learning. *Computer*, 51(5):42–49, 2018.

12. Guotai Wang, Wenqi Li, Maria A Zuluaga, Rosalind Pratt, Premal A Patel, Michael Aertsen, Tom Doel, Anna L David, Jan Deprest, Sébastien Ourselin, et al. Interactive medical image segmentation using deep learning with image-specific fine tuning. *IEEE transactions on medical imaging*, 37(7):1562–1573, 2018.

13. Theodora S Brisimi, Ruidi Chen, Theofanie Mela, Alex Olshevsky, Ioannis Ch Paschalidis, and Wei Shi. Federated learning of predictive models from federated electronic health records. *International journal of medical informatics*, 112:59–67, 2018.

14. Tsung-Ting Kuo, Hyeon-Eui Kim, and Lucila Ohno-Machado. Blockchain distributed ledger technologies for biomedical and health care applications. *Journal of the American Medical Informatics Association*, 24(6):1211–1220, 2017.

15. Edoardo Gaetani, Leonardo Aniello, Roberto Baldoni, Federico Lombardi, Andrea Margheri, and Vladimiro Sassone. Blockchain-based database to ensure data integrity in cloud computing environments. 2017.

16. Oded Goldreich. Secure multi-party computation. *Manuscript. Preliminary version*, 78, 1998.

17. Diego Ongaro and John K Ousterhout. In search of an understandable consensus algorithm. In *USENIX Annual Technical Conference*, pages 305–319, 2014.

18. Dragi Kimovski, Roland Mathá, Josef Hammer, Narges Mehran, Hermann Hellwagner, and Radu Prodan. Cloud, fog or edge: Where to compute? *IEEE Internet Computing*, 2021.

19. Andreas Leibetseder, Sabrina Kletz, Klaus Schoeffmann, Simon Keckstein, and Jörg Keckstein. GLENDA: gynecologic laparoscopy endometriosis dataset. In *MultiMedia Modeling - 26th International Conference, MMM 2020, Daejeon, South Korea, January 5-8, 2020, Proceedings, Part II*, volume 11962 of *Lecture Notes in Computer Science*, pages 439–450. Springer, 2020.

Dragi Kimovski is a tenure-track researcher at the Institute of Information Technology (ITEC), Klagenfurt University. He earned his Ph.D. in 2013 from the Technical University of Sofia (Bulgaria). He was an assistant professor at the University for Information Science and Technology in Ohrid (North Macedonia) and a senior researcher at the University of Innsbruck (Austria). His research interests include fog and edge computing, multi-objective optimization, and distributed storage.

Sasko Ristov is a postdoctoral university assistant at the University of Innsbruck, Austria. His research interests include performance modeling and optimization of parallel and distributed systems, particularly workflow applications and serverless computing. Dr. Ristov has a Ph.D. degree in computer science from Ss. Cyril and Methodius University, Skopje, North Macedonia, where he was Assistant Professor (2013-2017).

Radu Prodan is professor in distributed systems at ITEC, Klagenfurt University. He received his Ph.D. degree in 2004 from the Vienna University of Technology and was Associate Professor until 2018 at the University of Innsbruck (Austria). His research interests include performance and resource management tools for parallel and distributed systems.