Blockchain and Smart Contracts for Provenance of Deep Learning Content in Healthcare

Yuri Gordienko*, Oleg Alienin, Oleksandr Rokovyi, Volodymyr Valko and Sergii Stirenko
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute",
Kyiv, Ukraine
*yuri.gordienko@gmail.com

Gilles Fedak**, Oleg Lodygensky***
iExec Blockchain Tech
Lyon, France
**gilles.fedak@iex.ec
***oleg.lodygensky@iex.ec

***Abstract*—Due to advances of artificial intelligence (AI) and deep learning (DL) techniques, the opportunities for the reliable medical classification and prediction of some diseases become possible in recent years. Some predictions made by DL neural network trained on the huge medical datasets (MD) sometimes overcome the experts in the field and DL-models can be considered as useful AI-screening tools and good assistant for the real doctors. In this context, the proof of authenticity (PoA) of such deep learning content (DLC) (like datasets, models, etc.) is very important to realize the origin and evolution of DLC. At the moment there are no convenient solutions that can provide history tracking and provenance of DLC. In this paper, we provide a general framework using Ethereum smart contracts to track back the provenance and evolution of DLC to its original source even if the DLC was edited (e.g. DL models were re-trained or/and datasets were updated) by anonymous authors. The main principle behind the solution is that if the DLC can be credibly traced to a trusted or reputable source, the DLC can then be real and authentic. The solution is proposed in the healthcare context and for medical DLC, but it can be applied to any other form of DLC.**

***Index Terms*—blockchain, Ethereum, smart contracts, healthcare, artificial intelligence, deep learning, convolutional network, lung disease, pneumonia, tuberculosis, COVID-2019.**

I. Introduction

Due to the fast development of the new information technologies (ITs) healthcare organizations can provide the better services, but it is related with the challenges of scattered big data, complex data processing tools, reliability of computer-aided diagnostics (CADe/CADx) systems, confidentiality and privacy of patient electronic health record (EHR) data, and others. The combination of emerging ITs such as artificial intelligence (including deep learning), decentralized infrastructures (including blockchain) and distributed computing (including cloud computing) allow "consumers" (i.e. healthcare stakeholders like patients, doctors, hospitals, research labs, companies, etc.) to unlock the potential of ITs to obtain significant economic and social returns.

Now, it is crucial to have techniques to detect the reliable medical deep learning content (DLC) that include medical datasets, medical meta-information (professional expert labeling), models trained on these medical datasets with related medical meta-information, and so on. Achieving this purpose is not difficult if there is a credible, secure, and trusted way to trace the history of DLC. Users should be given access to a trusted data provenance of the reliable DLC, and be able to track back an item in history to prove its originality and authenticity. This mechanism can help assist consumers from being tricked or lured into believing in unreliable or fake DLC.

Current solutions are available to prove the authenticity of DLC due to the open-source repositories (like github and others) and databases (like Kaggle and others). The only approach currently used consists in the substantial manual work related with checking the source of DLC at these repositories or databases. But there are no established methods for checking the originality of an proprietary or open source DLC. It is extremely difficult to determine in a credible and trusted way the true origin of a proposed DLC. A typical consumer usually uses online search engines to try to find relevant sources of origin (posts, blogs or reviews) on the DLC to judge its authenticity. Hence, there is an immense need for a Proof of Authenticity (PoA) system for online DLC to identify trusted published sources and therefore be able to combat fake/unreliable models, medical datasets, meta-information, etc. Blockchain has the ability to provide immutable data and transactions in a decentralized distributed ledger [1]. Blockchain applications now are numerous, and the blockhain technology potentially and actually can disrupt many industries, for example, digital media [2], health management [4], IoT [5], AI/DL [3], etc.

Blockchain has capabilities to provide key features that can be utilized for proving authenticity and originality of digital assets in a way that is decentralized, highly trusted and secure, with tamperproof records, logs, and transactions which are openly accessible to all in case of permissionless blockchain, or restricted to certain participants in case of permissioned blockchain. For the reliable medical DLC, the permissionless or public blockchain is the most suitable. We base our solution in this paper on the public Ethereum blockchain with smart contracts to govern and capture the history of transactions made to the medical DLC.

In this paper, we propose a blockchain-based solution and a generic framework for the proof of authenticity of the medical DLC that may include medical datasets, medical meta-information (professional expert labeling), models trained on these medical datasets with related medical meta-information, medical predictions (results of application of the medical models on the new medical data), etc. Our solution allows for publicly accessible, trusted, and credible data provenance, with tracking and tracing history of a published medical DLC. Our solution focuses on medical DLC, but the solution framework provided in this paper is generic enough and can be applied to any other form of DLC as datasets, meta-information, models trained on these datasets with related meta-information, etc.

The paper contains Section II with description of the related work, Section III with the general view on the proposed blockchain-based solution, Section IV with some DLC specific implementation details, Section V with discussion of the solution, and Section VI with the final conclusions.

II. State of the Art and Related Work

A. Deep Learning in Healthcare Context

Deep Learning (DL), which is considered as a subset of Artificial Intelligence (AI), has demonstrated significant successes during the last years supported by rapid evolution of computational power and the wide availability of massive new open datasets [6]. Recently, DL has proved to be effective in many academic and industrial areas, including medical applications, such as segmentation of anatomical structures, various abnormality (like tumor) detection, disease classification, computer-aided diagnosis, etc [7]. Healthcare become especially successful field of DL applications due to progress of medical hardware generating the huge volume of data (for example, up to $10^{1}8$ bytes in USA only according to the recent reports [8]) and storage systems [9]. And the promising results of medical imaging were demonstrated for computer-aided diagnostics (CADe/CADx) in dermatology, radiology, ophthalmology, and others [9]. DL-based CADe/CADx systems could play the assistive role for medical experts by offering second opinions and attracting attention to some suspicious areas in images. Moreover, sometimes DL-based CADe/CADx systems can overcome the human experts in some complex diagnostics procedures. Some examples of the a deep learning-based computer-aided diagnosis system can be found on the market, for example, "Dr. Pecker", which is a medical image analysis software product [10].

B. Blockchain in Healthcare Context

Blockchain has a range of built-in features (like distributed ledger, decentralized storage, authentication, security, and immutability) and now has a range practical applications in many industries, including healthcare. Blockchain applications in the healthcare sector generally face harder authentication, interoperability, and record sharing requirements, for example, stated in the legal requirements, such as Health Insurance Portability and Accountability Act of 1996 (HIPAA) [11]. Currently blockchain is proposed to be utilized in numerous healthcare applications for data sharing, access control, health records, managing an audit trail, supply chain [12], [13], [14]. The roadmap for a blockchain-assisted decentralized bionetwork of private healthcare data was proposed to use new methodologies to drug discovery and precautionary healthcare [15]. Several existing specific software solutions with their detailed analysis can be found [14] and briefly outlined below: Gem Network is designed to share health data through decentralized network with legal issues addressed [16], OmniPHR proposes the distributed architecture model for scaling up during sharing the patient records [17], MedRec attempts to handle the mining incentive problem for sharing the health-related data [4], [18], pervasive social network (PSN) helps create the system that handles security of IoT devices [19], Virtual Resources allow to make persistent data storage that is safe, secure, and scalable [20], context-driven data logging approach is focused on adding a level of confidence to data logging [21], MeDShare aimed on security and data authentication for sharing the medical data [22], the trial and precision medicine platform designed for adding integrity and data access security to medical data [23], Healthcare Data Gateways application proposes to use the decentralized nature of blockchain technology to add a level of security and data integrity while simplifying legal issues [24], a new model of storing health data using IPFS based off-chain storage, so this model more decentralized and does not rely on any third-party providers [25].

C. Deep Learning and Blockchain Integration

Immediately after its appearance, blockchain became an extremely popular technology in many industries. But, because blockchain is highly cost effective in eliminating the need for a centralized authority to govern and verify interactions and transactions among several participants and also creates a secure, synchronized and shared timestamped records that cannot be altered it makes effective combination with AI technology that needs to address these issues in order to be even more efficient and adaptable for use in areas that require confidentiality and decentralization such as medicine [26].

Numerous works were dedicated to combining AI and blockchain to help solve medical problems. For example, attempts to integrate blockchain technology with artificial intelligence were made for cardiovascular medicine [28]. The new architecture for building large distributed medical data sets from various distributed resources and different data sources was proposed to provide the data (given by hospitals, individual patients or service providers) for DL researches in medical domain [27]. the concept of framework for AI-blockchain system for EHR management system was proposed [29], the specific EHR management system for university hospital with additional usage of neural networks for diagnose skin, heart and other diseases to facilitate doctors to make a diagnosis was considered [30]. The representation of patients' entire raw EHR records based on the Fast Healthcare Interoperability Resources (FHIR) format and deep learning methods using this representation were proposed to drive personalized medicine and improve healthcare quality [31].

This work is dedicated to the problem of the medical DLC provenance which intrinsically appear in the context of CADe/CADx approaches, when medical personnel should use numerous medical DLC components and be sure about reliability of their origin and evolution. This aspect is especially important in the view of the appearance of automated machine and deep learning approaches that allow to create the very useful medical DLC in automated way by independent and often trustless subsidiaries. This problem did not obtained enough attention yet, despite several valuable works dedicated to the similar aspects, but in another contexts. For instance, for the video deepfake problem, the blockchain-based framework is proposed for undisputed traceability to the original source [2]. In another research, the decentralized transfer learning platform is proposed to allow sharing of DL data and expertise and the sharing mechanism is based on smart contracts to ensure the intellectual property of an individual remains protected [3]. That is why this work was inspired by these approaches [2], [3] and dedicated to the problem of the DLC provenance in healthcare context by means of blockchain technologies.

III. General Structure of the Framework

The proposed framework is based on Ethereum blockchain's basic principles of transparency and traceability where secure and trusted history tracking and tracing are provided in a decentralized manner with no intermediaries or trusted third parties. This section describes and details our blockchain-based approach for proof of authenticity (PoA) of medical DLC. This approach can also be used for other types of DLC that can contain any kind of multimedia information from other than medical applications which will be described in details elsewhere [32].

From the general point of view, the proposed blockchain-based framework has the following key components (Fig. 1):

- Medical Dataset (D_i): It can contain the important medical information as a raw medical data (RD_i) (results of any kind of medical examination) and the related medical metadata or medical labels (L_i) (Fig. 1). The latter contains information about the actual RD_i: related to the medical device providing the RD_i, capture settings, date and time of capture, as well as logs and manually added information that the D_i "author" (creator) can add. Every medical dataset D_i can be associated with an Ethereum smart contract that can be created by an author or consumer on the basis of labeling the raw data, i.e. $D_i = L(L_i, RD_i)$, where L is some function depicting raw data labeling process (see below, e.g. in Fig. 4). The Ethereum address of the author as well as the address of the smart contract are integral parts of the D_i.
- Medical Trained DL Model (M_i): It can contain the raw model RM_i (i.e the description of a DL model architecture), the related model metadata about the training dataset and training process (number of epochs, learning rate schedule, etc.), and model weights trained on the concrete D_i. The raw model RM_i contains information about the actual DL network used: layers, activation functions, etc. The M_i reflects the trained state of RM_i.

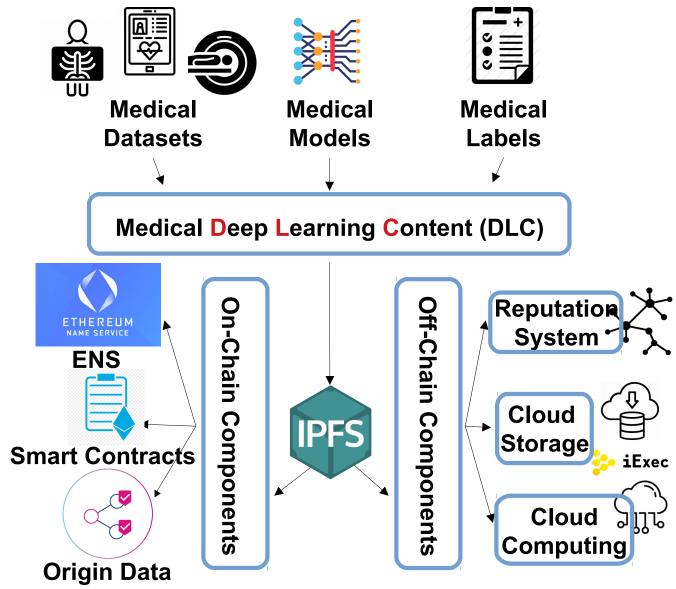

Fig. 1. General structure of the proposed blockchain-based framework.

Every model M_i will be associated with an Ethereum smart contract that can be created by an "author" (DL model developer here). The Ethereum address of the author as well as the address of the smart contract are integral parts of the model M_i.
- Medical Prediction (P_i) or predicted label: It can be result of application of some medical DL model M_i to some new raw data RD_i with the predicted label (prediction of its state) P_i. It seems to be that P_i results in creation of the "new" data D_j, because P_i contains the new raw data RD_i and the newly created L_i (labeling as a result of prediction). But actually P_i cannot become the "new" data D_j, except for human labeling L_i^H (see below, e.g. in Fig. 4) or its proper characterization by human experts. Every P_i will be associated with an Ethereum smart contract that can be created by an author or consumer. The Ethereum address of the author as well as the address of the smart contract are integral parts of the P_i.
- IPFS Storage: The DLC are stored on a decentralized, content-addressable, peer to peer file system such as the InterPlanetary File System (IPFS) [33]. IPFS generates a unique hash which is the address of a bundle of files containing the DLC.
- On-chain Components: After the DLC's IPFS hash is created, a smart contract is created by the original author (owner) on the Ethereum blockchain. The contract has attributes and variables to capture the DLC details and author's information. Variables are also used to store static information such as the DLC related data as well as the contract state. In its turn, any edited DLC by a secondary author will have its own smart contract with a link to the original DLC. Hence, all edited DLCs of any original DLC are "child" DLCs and are available in a list

in the original DLC's smart contract. Therefore, a user who would like to trace a DLC to its origin can easily do so using the on-chain components such as the smart contract which has the list of all the children DLC's smart contracts as well as a link to their parent's smart contract.

- ENS: A user can also make use of the Ethereum Name Service (ENS) [34] to associate Ethereum address of authors/consumers to a human-readable texts capturing the author/consumer's real identity including name, organization, and profile.
- Off-chain Components: A user tracing the data can also look at the off-chain components which are unavoidable part of the proposed framework, for example, cloud computing resources [35], because of the huge size of DLC and related computations.

IV. Implementation Details

A. DLC Developer View

In Fig. 2 the developer view on DLC evolution is shown. Several possible evolution tracks are shown for creation of the new versions of datasets and DL models.

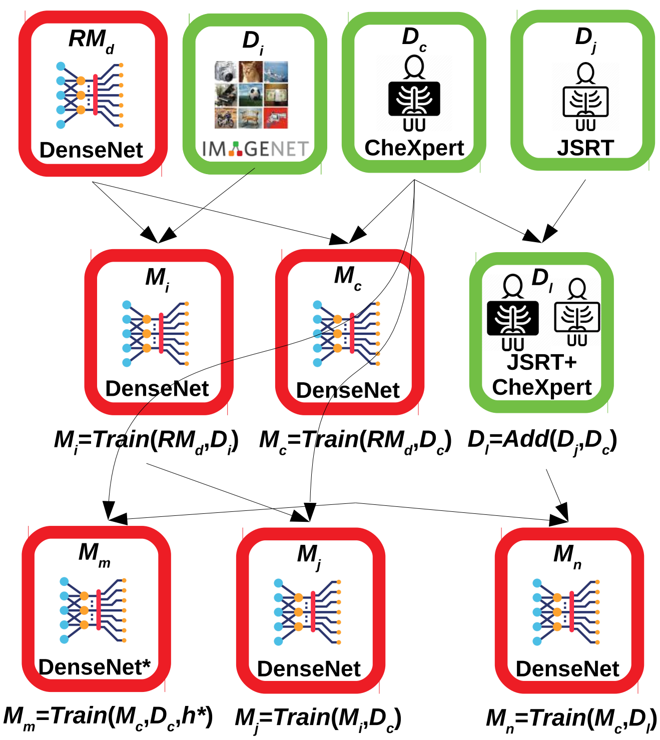

Fig. 2. Possible evoluton tracks for creation of the new versions of datasets and DL models.

Under the current conditions and without blockchain involvement the typical tracks can include:

- training the raw model RM_d (for example, from DenseNet family [37]) on the medical dataset D_c (for example, CheXpert [38]) could lead to the new trained model $M_c = Train(RM_d, D_c)$, and then its additional re-training on the same dataset D_c (CheXpert [38]), but with other hyperparameters (h^*) could result in the new trained model $M_m = Train(M_c, D_c, h^*)$ where $Train$ is the function depicting re-training process (the left rectangle M_m in the bottom of Fig. 2);
- training the raw model RM_d (for example, from DenseNet family [37]) on the standard general-purpose dataset D_i (for example, Imagenet [36]) could lead to the new trained model $M_i = Train(RM_d, D_i)$, and then its additional re-training on the other dataset D_e (for example, CheXpert [38]) could result in the new trained model $M_i = Train(M_i, D_e)$ (the central rectangle M_j in the bottom of Fig. 2);
- combining the medical dataset D_e (for example, CheXpert [38]) with other medical dataset D_e (for example, JSRT [39]) could lead to the new combined dataset $D_i = Add(D_j, D_e)$ where Add is the function depicting data combining process (Fig. 2), and then additional re-training of model M_c on the combined dataset D_i could result in the new trained model $M_n = Train(M_c, D_i)$ (the right rectangle M_m in the bottom of Fig. 2).

B. Tracing a DLC to its Origin

Usually, DLC developers clearly understand the evolution and origin of their DLC final versions of combined datasets (like D_i in Fig. 2) or re-trained models (like M_m, M_j, M_n in Fig. 2). But from the consumer point of view it is not so evident, especially in the view of the numerous new versions of DLC for CADe/CADx of various diseases (like cancer [40]–[44], tuberculosis [45]–[47], several other lung abnormalities [38], [48], [49], etc.) published by the scientific community (numerous scientists and independent developers). It has become evident during the current COVID-2019 pandemia when abundant volume of the related DLC content had appeared and has been reviewed thoroughly and criticized [50].

The main aim of the proposed solution is to assist consumers in tracing back the DLC with multiple versions to its origin. If the DLC cannot be traced reliably to its original publisher, then it cannot be trusted, especially in the healthcare context. Usually, consumers have access to some DLC provided by developers with the limited metadata about the real origin of the DLC. The matter is that the DLC is not only the function of its previous states, but also the function of the processes (like training with specific hyperparameters or combining datasets with some specific pre-processing) leading to these previous states. That is why the blockchain based methods (including smart contracts, IPFS, ENS, and other on- and off-chain components to establish authenticity of the DLC) should be used for the reliable provenance tracking especially for the trustless scientific community. Fig. 3 shows how a consumer could track the DLC orgin and evolution due to involvement of smart contracts. For example, contract A corresponds to the raw model RM_d trained on the dataset D_i (with the contract B) with the resulting model $M_i = Train(RM_d, D_i)$ (with the contract E which is the child of its parent contracts A and B). The further evolution of DLC consists in re-training the model

M_i (with the contract E) on the dataset D_c (with the contract C) leading to the resulting model $M_j = Train(M_i, D_c)$ (with the contract G which is the child of its parent contracts E and C) (the red tracks in Fig. 3). As to the DLC of dataset type, combining the medical dataset D_c (with the contract C) with other medical dataset D_j (with the contract D) leads to the resulting dataset $D_l = Add(D_j, D_c)$ (with the contract F which is the child of its parent contracts C and D) (the green track in Fig. 3).

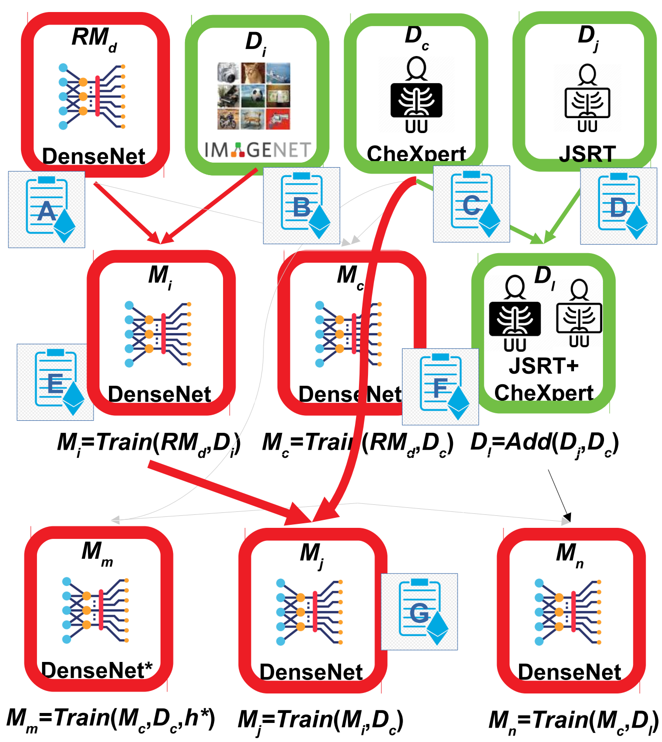

Fig. 3. DLC evolution for model (red) and dataset (green) tracks.

C. DLC consumer View

In Fig. 4 the consumer (medical personnel) view on DLC is shown, where the medical DLC (like re-trained medical models like M_m, M_j, M_n and the medical datasets D_e^{GT}, D_e^{NLP}) are assumed to be obtained from developers by the tracks shown in Fig. 3 with the correspondent blockchain contracts involved.

It should be noted that evolution of the raw medical data (like RD_b) to the medical datasets can follow by different tracks depending on the nature of labeling process. In the case of the human labeling L_a^H (or ground truth (GT) labeling) for the raw medical data RD_b by the human expert E_e^{Human} one can obtain the dataset $D_e^{GT} = L(L_a^H, RD_b)$. But in many cases computer-aided labeling L_a^{NLP} for the raw medical data RD_b is made by the natural language processing (NLP) DL models (like M_a^{NLP}) [51] with the resulting dataset $D_e^{NLP} = L(L_a^{NLP}, RD_b)$. For the data provenance both of these tracks could traced back by their contracts, for example, the contract CD_d^{GT} is the child of its parent contracts CRD_b and CL_a^H where the latter is the child of its parent contract CE_c^H. The consumers could obtain DLC (models like M_m, M_j, M_n with reliable provenance for constructing the following predictions $P_m = P(M_m, RD_b)$, $P_n = P(M_n, RD_b)$, $P_j = P(M_j, RD_b)$ (the violet region in Fig. 4). On the basis of this predictions, consumers can obtain the assistive CADe/CADx decision (by auction, voting, or other blockchain method) as to the availability of diseases/abnormality precursors in the raw medical data RD_b. But if these raw data RD_b have the proper GT labeling by human experts and actually provide the dataset D_d^{GT}, then metrics $A_m = Test(M_m, D_e^{GT})$, $A_n = Test(M_n, D_e^{GT})$, $A_j = Test(M_j, D_e^{GT})$ for these models M_m, M_j, M_n can be calculated after tests against the dataset D_d^{GT} (orange tracks in Fig. 4). Finally, these metrics can be used for the construction of the more effective ensembles of the medical DL models.

A front-end decentralized application (or consumer DApp) can be developed for the user to automate the authenticity process, or it can be integrated within DL frameworks or web-based services or applications to indicate authenticity of the used DLC. In Fig. 3, every DLC in the red/green tracks is associated with a smart contract that points to its parent DLC and every parent DLC is linked to its child, in a hierarchical fashion. As shown in Fig. 3, a consumer can trace smart contract G to its parent smart contracts C and E where the latter is traceable to the linked smart contracts A and B. The smart contract F can be traced back to the linked smart contracts C and D. These provenance data are openly accessible and available due to the Ethereum ledger.

V. Discussion

The depicted blueprint of the provenance data framework for the medical DLC does not describe all possible developer and consumer roles and tracks. For example, the additional contracts will be necessary to follow the origin of raw data with regard to the medical devices where they were captured that will demand taking into consideration some aspects of blockchain for IoT and Edge Computing [5], [19]. The off-chain cloud resources will be necessary for heavy computations related with training, predicting, and testing the DL models, and also large storage cloud resources will be necessary for storing/pre-processing huge medical datasets. As a future work, we are in the process of developing front-end Dapps on the basis of iExec SDK [35] for users to automate the establishment of proof of authenticity of published DLCs, which could be especially promising with regard to the necessity to perform the heavy off-chain computations by decentralized cloud computing paradigm. The proposed blockchain-based solution can ensure the key aspects of decentralized DLC framework like integrity, accountability, authorization, availability and non-repudiation. All transaction history as well as the provenance data available for the consumers to track and trace a DLC to its origin are tamper proof. Every participating entity is accountable for its actions on the ledger. All transactions taking place on the blockchain network are cryptographically signed by the initiator and no one can deny

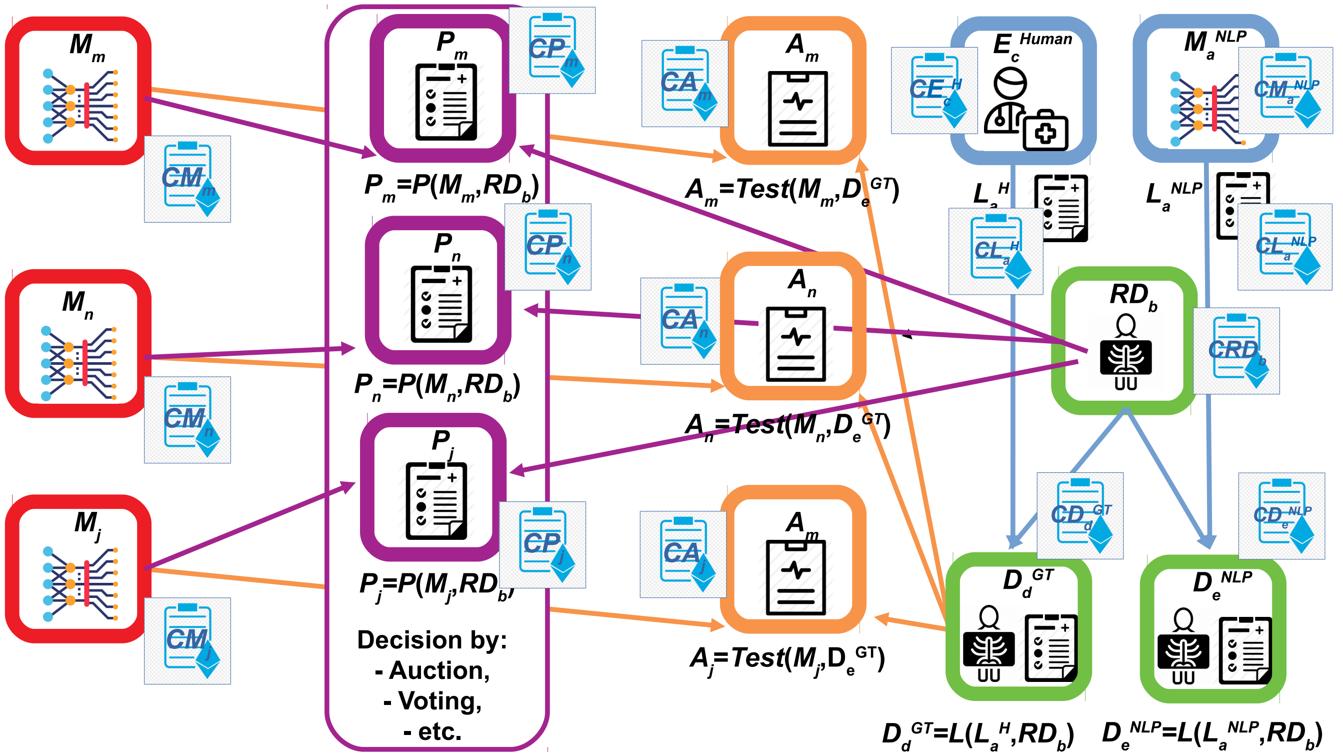

Fig. 4. The consumer view: predictions P_m, P_n, P_j (violet) and their metrics after tests A_m, A_n, A_j (orange) for the new data D_d^{GT} (or D_e^{NLP}) obtained after labeling (blue) the new raw data RD_b by human experts E_e^{Human} (or by NLP models M_a^{NLP}).

their own actions. The participants can always access the smart contracts once deployed to the blockchain network. The information stored on the ledger is saved in a distributed and decentralized way and is not subject to hacking, compromise or being a single point of failure. The relevant analysis of the most interesting use cases (like shown in Fig. 4) with examples of Solidity smart contracts, a pluggable DApp component and estimation of the related operational costs will be described in details elsewhere [32].

VI. Conclusion

In this paper, we have presented a blockchain-based solution to search back the history of the medical DLCs in which a secure and trusted traceability to the original DLC creator or source can be established, in a decentralized manner. The proposed solution could use a decentralized storage system IPFS, Ethereum name service, and off-chain cloud component support by iExec platform. At the moment the proposed blueprint of framework, system design, and implementation details are generic enough that can be applied not only for the medical DLC, but also to other types of DLC such as various unstructured (images, sequences, video, audios, etc.) and structured data. The solution can help to resist fake or unreliable medical DLCs by helping consumers to determine if a medical DLC is traceable to a trusted and reputable source. If a medical DLC is not traceable, then the medical DLC cannot be trusted. Moreover, the smart contract-based solution can provide a trusted way for secondary authors to request permission from the original author to copy and edit DLCs. The future research is dedicated to development of a pluggable DApp component and estimation of the related operational costs in terms of Ether and Gas when the smart contract will be deployed on the real Ethereum network with intensive use of off-chain components like cloud computing and storage resources.

References

[1] I. Bashir, “Mastering Blockchain: Distributed ledger technology, decentralization, and smart contracts explained,” Packt Publishing Ltd., 2018.
[2] H. R. Hasan and K. Salah, “Combating deepfake DLCs using blockchain and smart contracts,” IEEE Access, 7, 41596-41606, 2019.
[3] A. ul Haque, M. S. Ghani, and T. Mahmood, “Decentralized Transfer Learning using Blockchain & IPFS for Deep Learning,” In 2020 Int. Conf. on Information Networking (ICOIN) (pp. 170-177), IEEE, 2020.
[4] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, “Medrec: Using blockchain for medical data access and permission management,” in Proc. 2nd Int. Conf. Open Big Data (OBD), pp. 25–30, 2016.
[5] M. A. Khan and K. Salah, “IoT security: Review, blockchain solutions, and open challenges,” Future Gener. Comput. Syst., vol. 82, pp. 395–411, 2018.
[6] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature 521, 436–444, 2015.
[7] Y. W. Chen, and L. C. Jain, “Deep Learning in Healthcare,” Springer, 2020.
[8] Stanford Health, “Harnessing the power of data in health,” Stanford Medicine 2017 Health Trends Report, 2017.
[9] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, ... , and J.Dean, “A guide to deep learning in healthcare,” Nature Medicine, 25(1), 24-29, 2019.
[10] G. Cheng and L. He, “Dr. Pecker: A Deep Learning-Based Computer-Aided Diagnosis System,” Deep Learning in Healthcare: Paradigms and Applications, 171, 203, 2019.

[11] "Health insurance portability and accountability act of 1996," Public law, 104, 191, 1996.

[12] H. Rathore, A. Mohamed, and M. Guizani, "Blockchain applications for healthcare. In Energy Efficiency of Medical Devices and Healthcare Applications," (pp. 153-166), Academic Press, 2020.

[13] M. Holbl, M. Kompara, A. Kamisalic, and L. Nemec Zlatolas, "A systematic review of the use of blockchain in healthcare," Symmetry, 10(10), 470, 2018.

[14] T. McGhin, K. K. R. Choo, C. Z. Liu, and D. He, "Blockchain in healthcare applications: Research challenges and opportunities," J. Network and Computer Applications, 135, 62-75, 2019.

[15] V. B. Lobo, J. Analin, R. M. Laban, and S. S. More, "Convergence of Blockchain and Artificial Intelligence to Decentralize Healthcare Systems," In 2020 Fourth Int. Conf. on Computing Methodologies and Communication (ICCMC) (pp. 925-931), IEEE, 2020.

[16] M. Mettler, "Blockchain technology in healthcare: The revolution starts here," In 2016 IEEE 18th Int. Conf. on e-Health Networking, Applications and Services (Healthcom) (pp.1-3), IEEE, 2016.

[17] A. Roehrs, C. A. da Costa, and R. da Rosa Righi, "OmniPHR: A distributed architecture model to integrate personal health records," J. Biomedical Informatics, 71, 70-81, 2017.

[18] N. Nchinda, A. Cameron, K. Retzepi, and A. Lippman, "MedRec: A network for personal information distribution," In 2019 Int. Conf. on Computing, Networking and Communications, pp. 637-641, IEEE, 2019.

[19] J. Zhang, N. Xue, and X. Huang, "A secure system for pervasive social network-based healthcare," IEEE Access, 4, 9239-9250, 2016.

[20] M. Samaniego, and R. Deters, "Hosting virtual iot resources on edge-hosts with blockchain," In 2016 IEEE Int. Conf. on Computer and Information Technology (CIT) (pp. 116-119), IEEE, 2016.

[21] M. Siddiqi, S. T. All, and V. Sivaraman, "Secure lightweight context-driven data logging for bodyworn sensing devices," In 2017 5th Int. Symp. on Digital Forensic and Security, pp. 1-6, IEEE, 2017.

[22] Q. I. Xia, E. B. Sifah, K. O. Asamoah, J. Gao, X. Du, and M. Guizani, "MeDShare: Trustless medical data sharing among cloud service providers via blockchain," IEEE Access, 5, 14757-14767, 2017.

[23] Z. Shae, and J. J. Tsai, "On the design of a blockchain platform for clinical trial and precision medicine," In 2017 IEEE 37th Int. Conf. on Distributed Computing Systems (ICDCS) (pp. 1972-1980), IEEE, 2017.

[24] X. Yue, H. Wang, D. Jin, M. Li, and W. Jiang, "Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control," J. of Medical Systems, 40(10), 218, 2016.

[25] R. Kumar, N. Marchang, and R. Tripathi, "Distributed Off-Chain Storage of Patient Diagnostic Reports in Healthcare System Using IPFS and Blockchain," In 2020 Int. Conf. on COMmunication Systems & NETworkS (COMSNETS) (pp. 1-5), IEEE, 2020.

[26] K. Salah, M. H. U. Rehman, N. Nizamuddin, and A. Al-Fuqaha, "Blockchain for AI: Review and open research challenges," IEEE Access, 7, 10127-10149, 2019.

[27] Z. Shae, and J. Tsai, "Transform blockchain into distributed parallel computing architecture for precision medicine," In 2018 IEEE 38th Int. Conf. on Distributed Computing Systems (ICDCS) (pp. 1290-1299), IEEE, 2018.

[28] C. Krittanawong, A. J. Rogers, M. Aydar, E. Choi, K. W. Johnson, Z. Wang, and S. M. Narayan, "Integrating blockchain technology with artificial intelligence for cardiovascular medicine," Nature Reviews Cardiology, 17(1), 1-3, 2020.

[29] Y. Wehbe, M. Al Zaabi, and D. Svetinovic, "Blockchain ai framework for healthcare records management: Constrained goal model," In 2018 26th Telecommunications Forum (TELFOR) (pp. 420-425), IEEE, 2018.

[30] W. Cao, J. Su, Z. Peng, W. Xu, and Q. Liu, "Electronic Medical Record of University Hospital Based on Deep Learning," In IOP Conference Series: Materials Science and Engineering, Vol. 569, No. 5, p. 052110, IOP Publishing, 2019.

[31] A. Rajkomar, E. Oren, K. Chen, A. M. Dai, N. Hajaj, M. Hardt, ... and P. Sundberg, "Scalable and accurate deep learning with electronic health records," NPJ Digital Medicine, 1(1), 18, 2018.

[32] Yu.Gordienko et al, "Provenance Data Platform for Deep Learning Content," unpublished.

[33] Juan Benet and Protocol Labs. Accessed: June 30, 2020. [Online]. Available: https://github.com/ipfs/ipfs.

[34] Ethereum Name Service. Accessed: June 30, 2020. [Online]. Available: https://ens.domains/

[35] "Blockchain-Based Decentralized Cloud Computing," Whitepaper, Version 3.0, 2017 - 2018. Accessed: June 30, 2020. [Online]. Available: https://iex.ec/wp-content/uploads/pdf/iExec-WPv3.0-English.pdf.

[36] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," In 2009 IEEE Conf. on Computer Vision and Pattern Recognition, pp. 248-255, IEEE, 2009.

[37] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 4700-4708, 2017.

[38] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, ... and J. Seekins, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," In Proc. AAAI Conf. on Artificial Intelligence, Vol. 33, pp. 590-597, 2019.

[39] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "Development of a digital image database forchest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," AJR 174, 71-74, 2000.

[40] C. Liu, B. Wang, Q.Jiao, and M. Zhu, "Reducing False Positives for Lung Nodule Detection in Chest X-rays using Cascading CNN," In 2019 14th IEEE Conf. on Industrial Electronics and Applications (ICIEA), pp. 1204-1207, IEEE, 2019.

[41] Y. Gordienko, P. Gang, J. Hui, W. Zeng, Y. Kochura, O. Alienin, ... and S. Stirenko, "Deep learning with lung segmentation and bone shadow exclusion techniques for chest x-ray analysis of lung cancer," In Int. Conf. on Computer Science, Engineering and Education Applications, pp. 638-647, Springer, Cham, 2018.

[42] M. Negassi, R. Suarez-Ibarrola, S. Hein, A.Miernik, and A. Reiterer, "Application of artificial neural networks for automated analysis of cystoscopic images: a review of the current status and future prospects," World Journal of Urology, 1-10, 2020.

[43] Q. Hu, L. F. D. F. Souza, G. B. Holanda, S. S. Alves, F. H. D. S. Silva, T. Han, and P. P. Reboucas Filho, "An effective approach for CT lung segmentation using mask region-based convolutional neural networks," Artificial Intelligence in Medicine, 103, 101792, 2020.

[44] R. M. Sundhari, "Enhanced histogram equalization based nodule enhancement and neural network based detection for chest x-ray radiographs," J. Ambient Intelligence and Humanized Computing, 1-9, 2020.

[45] S. Stirenko, Y. Kochura, O. Alienin, O. Rokovyi, Y. Gordienko, P. Gang, and W. Zeng, "Chest X-ray analysis of tuberculosis by deep learning with segmentation and augmentation," In 2018 IEEE 38th Int. Conf. on Electronics and Nanotechnology (ELNANO), pp. 422-428, IEEE, 2018.

[46] E. Tasci, "Pre-processing Effects of the Tuberculosis Chest X-Ray Images on Pre-trained CNNs: An Investigation," In Int. Conf. on Artificial Intelligence and Applied Mathematics in Engineering, pp. 589-596, Springer, Cham, 2019.

[47] D. Verma, C. Bose, N. Tufchi, K. Pant, V. Tripathi, and A. Thapliyal, "An efficient framework for identification of Tuberculosis and Pneumonia in chest X-ray images using Neural Network," Procedia Computer Science, 171, 217-224, 2020.

[48] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, ... and M. P. Lungren, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," arXiv:1711.05225, 2017.

[49] P. Gang, W. Zeng, Y. Gordienko, Y. Kochura, O. Alienin, O. Rokovyi, and S. Stirenko, "Effect of Data Augmentation and Lung Mask Segmentation for Automated Chest Radiograph Interpretation of Some Lung Diseases," In Int. Conf. on Neural Information Processing, pp. 333-340, Springer, Cham, 2019.

[50] L. Wynants, B. Van Calster, M. M. Bonten, G. S. Collins, T. P. Debray, M. De Vos, ... and E. Schuit, "Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal," BMJ, 369, 2020.

[51] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 2097-2106, 2017.