DAYANANDA SAGAR UNIVERSITY
KUDLU GATE
BANGALORE – 560068
SCHOOL OF ENGINEERING

Blockchain for Deep Learning: Review and Open Challenges

Course Title: Introduction to Blockchain and distributed ledger

Course Code: 19CS3618

BACHELOR OF TECHNOLOGY
IN
COMPUTER SCIENCE & ENGINEERING

Submitted by
TEJAAL M- ENG19CS0334
6th SEMESTER
'F'-SECTION

Under the supervision of
Mrs. Nandini K

INDEX

ABSTRACT

In this paper, we explore the importance of integrating blockchain technology with deep learning. We review the existing literature focused on the integration of blockchain with deep learning. We classify and categorize the literature by devising a thematic taxonomy based on seven parameters; namely, blockchain type, deep learning models, deep learning specific consensus protocols, application area, services, data types, and deployment goals. We provide insightful discussions on the state-of-the-art blockchain-based deep learning frameworks by highlighting their strengths and weaknesses. Furthermore, we compare the existing blockchain-based deep learning frameworks based on four parameters such as blockchain type, consensus protocol, deep learning method, and dataset. Finally, we present important research challenges which need to be addressed to develop highly efficient, robust, and secure deep learning frameworks.

INTRODUCTION

The integration of blockchain with deep learning can bring several benefits, e.g., automated and trusted decision making, efficient data market management, data security, better model building for prediction purposes, model sharing, and enhancement of the robustness of the deep learning-based systems.

Blockchain is a revolutionary technology that maintains a shared ledger of data among participants in a decentralized network. It ensures that all ledger copies maintained by the participants are verified and proven consistent .

By design, blockchain represents a tamper-proof and tamper-resilient technology that assists in tracking the data to ensure that it has not been tampered with since its creation.The main advantages of novel blockchain technologies include data immutability, transparency, security, provenance, traceability, and operational visibility, and such benefits are envisioned by the decentralized and Peer-to-Peer (P2P) architecture of blockchain.

In summary, the key contributions of this paper are as follows:

• We devise a taxonomy to categorize and classify the existing literature related to blockchain-based deep learning frameworks based on seven important parameters.

• We present insights into the state-of-the-art blockchain-based deep learning frameworks by highlighting their strengths and weaknesses.

• We compare the blockchain-based deep learning frameworks based on important parameters.

• We discuss several research challenges that can affect the performance, accuracy, and prediction quality of existing blockchain-based deep learning frameworks.

BACKGROUND

A. Blockchain Technology

Blockchain technology can assist in protecting the Electronic health records (EHR) and Personal health records (PHR) of patients. It can ensure that the data is controlled and managed by patients, and it can be shared with other users in compliance with the patient's consent management policy. The consent management policies are implemented through self-executing smart contracts. However, blockchain becomes a costly technology if appropriate techniques are not implemented to manage large sized healthcare data. To take full advantage of blockchain in many healthcare applications, the pointers and linkers can play a valuable role to minimize the data size. In addition, decentralized storage systems are also capable to securely store large-sized data and avoid single-point-of-failure-related problems. The examples of most widely used decentralized storage systems in healthcare sector include InterPlanetary File System (IPFS) ,Cassandra, SWARM , Stor, OrbitDB, and Skeps, to name a few.

B. Deep learning

In the case of a deep learning algorithm, the models are intelligent and automatically extract the high-level latent space features from the basic form of the data. By design, the deep learning models consist of multiple layers. The lower-level layers are responsible for extracting lower-level features, while higher layers extract more abstract features from the input data.

C. Blockchain-based Deep Learning

a. Blockchain-based Deep Learning
b. Automatic Decision Making
c. Cumulative Judgements
d. Enhanced Robustness

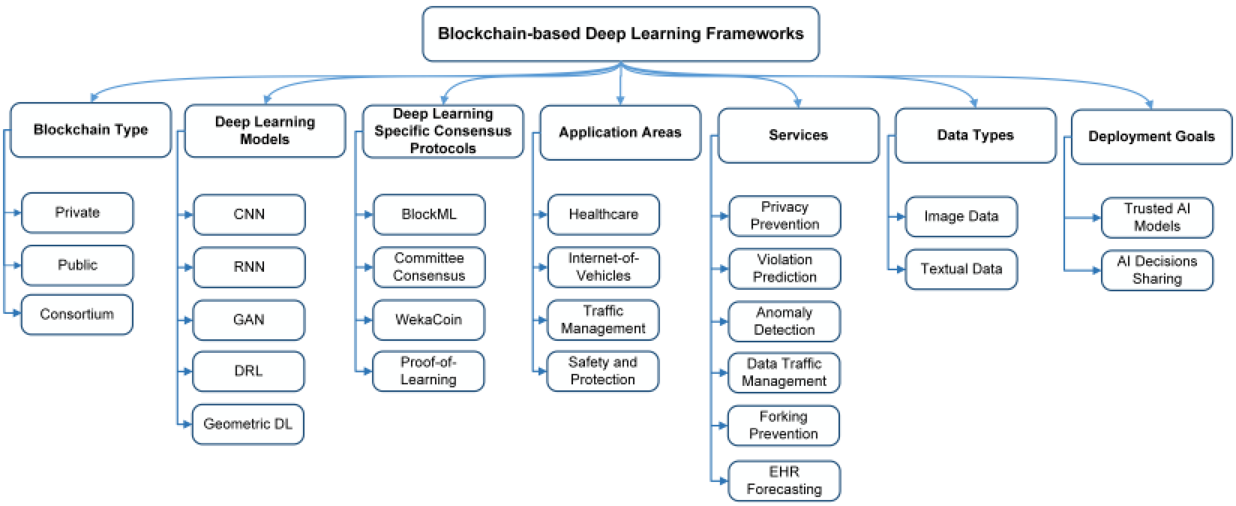

Fig. 4: A taxonomy of blockchain for deep learning frameworks.

TAXONOMY OF BLOCKCHAIN-BASED DEEP LEARNING FRAMEWORK

A. Blockchain Type

a. Public Blockchain:

Users access the ledger copy that is distributed among all nodes within the public blockchain network and performs transactions.

b. Private Blockchain:

The private blockchain platforms leveraged by the blockchain-assisted deep learning frameworks are controlled and managed by a single entity.

c. Consortium/Federated Blockchain:

A consortium blockchain functions as a permissioned network and, multiple heterogeneous groups can have the authorization role, unlike private networks where a single authority is responsible for controlling and managing the network .

B. Deep Learning Models

a. Convolution Neural Network

The advantage of selecting CNN in blockchain-based studies is the minimum preprocessing time required by the algorithm because of choosing adaptable filters to determine the characteristics of the image.

b. Recurrent Neural Network

Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) are upgraded versions of RNN to address the shortcoming of RNN and are widely employed for accurate forecasting.

c. Generative Adversarial Networks (GAN)

The generator is responsible for producing new examples, whereas the discriminator learns to classify the data as real or fake.

d. Deep Reinforcement Learning (DRL)

Intelligent agents take actions in an environment consisting of DRL models to learn.

e. Geometric Deep Learning

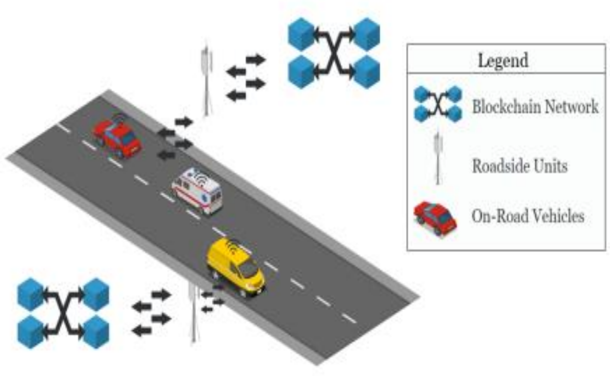

Fig. 5: Blockchain-based system for Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communication.

C. Deep Learning Specific Consensus Protocols

a. BlockML

b. Committee Consensus Mechanism

c. WekaCoin

d. **Proof-of-learning**

D. Application Areas

a. **Healthcare:**
b. **Internet of Vehicles**
c. **Traffic Management**
d. **Safety and Protection**

E. Services

a. **Privacy Preservation**
b. **Violation Prediction**

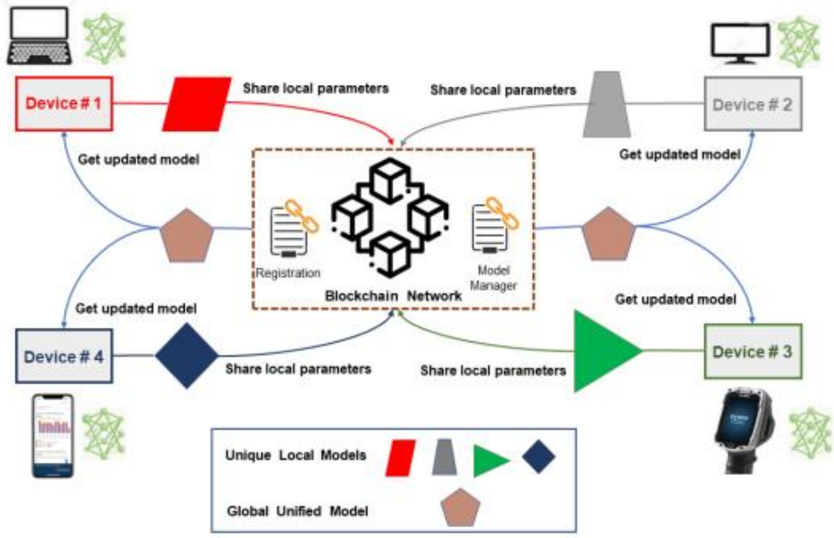

Fig. 6: Blockchain-based federated learning.

c. **Anomaly Detection**
d. **Data Traffic Management**
e. **Forking Prevention**
f. **EHR Forecasting:**

F. Data Types

a. **Image Data**
b. **Textual Data**

G. Deployment Goal

a. **Trusted AI Models**

The predicted results generated by the trusted AI models are highly reliable and correct. The model provenance enabled by blockchain technology can assist in verifying the trustworthiness of a deep learning model.

b. **AI Decisions Sharing**

The AI decision-sharing parameter describes the predictions made by the AI models based on time series data and secure sharing of the prediction among the potentially untrusted participants.

BLOCKCHAIN-BASED DEEP LEARNING FRAMEWORK

A. Review of Blockchain-based Deep Learning Frameworks

a. Data Analysis

The blockchain platform is used for sharing healthcare data, patient records, and ovarian cancer predictions made by the model with the participating organizations.

b. Provenance Data for AI Models

The main classes implemented by the proposed system to generate highly trusted AI models include participants, datasets, models, operations, and compute pipelines or projects. The record of immutable transactions stored on the blockchain has assisted the users in successfully maintaining the data provenance of the generated AI models.

c. Model Prediction

The study has proposed a novel consensus algorithm called proof of information , which is well suited for healthcare model prediction. The proposed study has employed an incremental federated learning model for implementation purposes.

d. Data Filtration

It was noticed that the majority of healthcare data comes from wearable devices at a continuous dynamic rate; such datasets can be highly valuable to build disease prediction models. The proposed study has mainly focused on dynamic data as this type of data is easily accessible via wearable devices.

e. Disease Classification

The data gathered after retraining can be stored in the hardware and can be accessed by using the pointers that are stored in a time-stamped ledger.

f. Combined Cooperative Positioning

The data gathered after retraining can be stored in the hardware and can be accessed by using the pointers that are stored in a time-stamped ledger.

g. Crowdsensing

The data gathered after retraining can be stored in the hardware and can be accessed by using the pointers that are stored in a time-stamped ledger.

h. Errors Reporting

i. Crowdsourcing

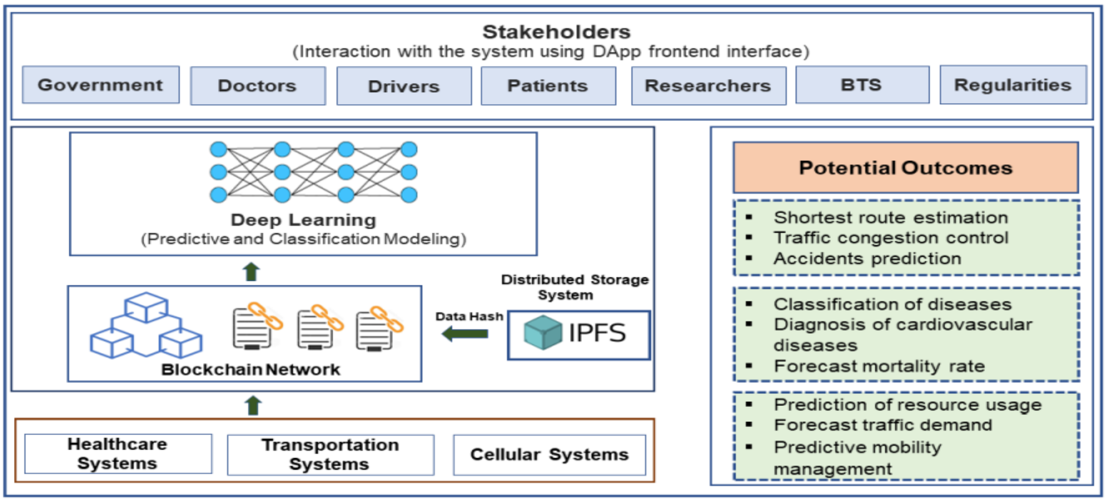

Fig: Highlighting the system components and participants and describing the outcomes resulting from the integration of deep learning with blockchain technology in various fields

j. Traffic flow Prediction

The researchers utilized this technology along with Gated Recurrent Unit (GRU) neural network to process live traffic data for improved traffic management. A consortium blockchain is used to keep the distributed ledger running by employing a small number of preselected miners.

k. Computations Offloading

TABLE II: A comparison of state-of-the-art blockchain-based deep learning frameworks. '-' represents that the corresponding information is not provided.

Category	Blockchain Type	Consensus Protocol	Deep Learning Method	Dataset
Ovarian Cancer Prediction [76]	-	-	One-shot Learning	Human Protein Atlas
Data Exchange [60]	Private	Proof-of-Information	Incremental Learning	-
EHR Prediction [84]	-	-	LSTM	EHR-based Dataset
Arrhythmia Classification [78]	-	-	SDA+ Sigmoid	MIT-BIH Database
Miner Node Selection [85]	-	Zero-Knowledge Proof	Deep Boltzmann Machine	-
Communication Security [86]	Private		DNN + Reinforcement Learning	-
Securing Blockchain [87]	-	Proof-of-Work	-	Game Theory Based Utility Function
Traffic Jam Prediction [82]	Public	Proof-of-Authority	ANN + LSTM	Historic Traffic Data + Custom Dataset
Traffic Flow Prediction [83]	Consortium	Delegated PBFT	GRU	-
Incident Prediction [72]	-	-	CNN	Custom Dataset
GPS Correction [81]	Public	Delegated PoS	DNN	Custom Dataset

l. Channel Protection

Singh et al. has proposed a blockchain and deep learning-based approach that has implemented Zero Knowledge Proof (ZKP) for validating the registered machines. The proposed method registers the machines such as a drone and verifies them using ZKP before the transactions from such machines can be granted. The proposed study selects a miner node using a novel selection algorithm that involves a deep Boltzmann machine.

m. Utility-based Blockchain Security

This utility function once fed into machine learning-based classification algorithms can assist in determining whether or not an intrusion is likely to occur.

B. Comparison of Existing Frameworks

The existing frameworks have implemented various services related to privacy prevention, violation prediction, anomaly detection, forking prevention, and EHR forecasting in healthcare, cellular data traffic, and vehicular communication networks. In a vehicular communication network, blockchain can be useful to securely share the decision made by an AI model such as road accident detection or traffic jamming with the authorized participating entities. Such data is useful to efficiently manage road traffic by evenly distributing the traffic on the roads . The data about the AI models stored on the blockchain includes training datasets, training results, model owner credentials, the source of data, and participants.

RESEARCH CHALLENGES AND OPPORTUNITIES

A. Platform Scalability

Considering the storage and computational requirements of the ever-growing blockchain ledger, the number of blocks and transactions to be added to the blockchain must be reduced considerably to satisfy the anticipated requests by the users. To carefully handle the scalability challenges of existing blockchain platforms, the compression algorithms having lightweight design, high compression ratio, and resource inexpensive nature should be integrated into the existing blockchain-assisted deep learning solutions.

B. Data Validity and Secure Sharing

A multi-layer blockchain architecture that supports data fusion and allows advanced analytical authentication for user groups can assist in securely sharing data between the participants.

C. Structural Enhancement and Storage Capacity

As the size of the blockchain increases, the efficiency of the network is profoundly affected. Deep learning approaches can be employed for compressing the data as well as assisting in minimizing the redundant data. Since data stored on the blockchain is permanent, hence, the growing size of the ledger is a big concern and needs to be addressed properly.

D. Platform Throughput and Latency

The high transaction execution latency is one of the major research challenges that should be addressed properly as it can affect the performance of processes that require quick decision-making.

E. Cryptocurrencies, Deep Learning-based Consensus Protocols, and Regulations

Currently, the generic blockchain consensus algorithms require time on the scale of seconds . Deep learning-specific consensus protocols can be designed based on the proofs related to the quality of the data, optimization techniques, nature of the learning models, 25 and total convergence time of the model. Lack of standards and regulatory frameworks is another challenge to the existing blockchain technology and it affects blockchain adaptability to deep learning frameworks.

F. High-Speed Computing/Storage Devices and Platform Interoperability

The interoperable blockchain platforms enable the participants of healthcare and vehicular communication networks to share data and information uninterruptably, securely, quickly, and seamlessly .The platform interoperability is affected by many factors such as the choice of blockchain-supported languages, consensus protocols, cryptographic hashing algorithms, and the type of data being used by the participants.

G. Secure Economical Models

Currently, existing deep learning models require high-performing computing devices for training purposes; and, the blockchain is an expensive storage medium. More research is needed to propose cost-efficient, resource-friendly, fast, and high-performance-based blockchain-assisted deep learning frameworks.

CONCLUSION

My concluding remarks along with the key recommendations include:

• Data traceability, immutability, and integrity features of blockchain technology can assist in identifying the volume and type of data collected to train deep learning models. However, the existing blockchain-based systems are incapable of efficiently handling data quality problems, particularly in the healthcare and transportation industries.

• The key performance metrics such as system throughput, execution latency, and block propagation time, data volume, conflicting interests of participants, and smart contract vulnerabilities can significantly impact the effectiveness of existing blockchain-based deep learning systems.

• Private blockchain platforms ensure data privacy through private channels and access control policies. However, public blockchain platforms are prone to data privacy leakage problems because they have a zero-access control policy. Public blockchain platforms can successfully capture the evolution of deep learning models as it progresses and records the model state during its creation, updating, or usage stages.

• The efficiency of blockchain-based applications is highly affected by increasing the size of the blockchain network. Deep learning approaches can be employed for compressing the data as well as minimizing the redundant data.

REFERENCES

- Blockchain for deep learning: review and open challenges
  - Muhammad Shafay, Raja Wasim Ahmad, Khaled Salah, Ibrar Yaqoob,Raja Jayaraman,Mohammed Omar
  - The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022
  - https://link.springer.com/article/10.1007/s10586-022-03582-7

- https://www.researchgate.net/publication/355566281_Blockchain_for_Deep_Learning_Review_and_Open_Challenges

- https://www.techrxiv.org/articles/preprint/Blockchain_for_Deep_Learning_Review_and_Open_Challenges/16823140

- https://ieeexplore.ieee.org/abstract/document/8598784%20/keywords#keywords