1 of 1

Download Print Save to PDF Save to list Create bibliography

Artificial Intelligence, Blockchain, Computing and Security: Volume 1 • Volume 1, Pages 690 - 698 • 1 January 2023

Document type
Book Chapter

Source type
Book

ISBN
978-100384581-2, 978-103249393-0

DOI
10.1201/9781003393580-102

View more

Blockchain-aware federated anomaly detection scheme for multivariate data

Selvakumar V.[a] ; Maaliw, Renato R.[b] ; Sharma, Ravi Mohan[c] ; Oak, Rajvardhan[d] ; Singh, Pavitar Parkash[e] ; Kumar, Ashok[f]

Save all to author list

[a] Department of Maths and Statistics, Bhavan's Vivekananda College of Science, Humanities and Commerce, Telangana, Hyderabad, India

[b] College of Engineering, Southern Luzon State University, Lucban, Quezon, Philippines

[c] Department of Computer Science and Applications, Makhanalal Chaturvedi National University of Journalism and Communication, Madhya Pradesh, Bhopal, India

[d] Department of Computer Science, University of California Davis, United States

[e] Mittal School of Business, Lovely Professional University, Punjab, Phagwara, India

[f] Department of Computer Science, Banasthali Vidyapith, Rajasthan, India

Hide additional affiliations

Full text options Export

Cited by 0 documents

Inform me when this document is cited in Scopus:

Set citation alert >

Related documents

Find more related documents in Scopus based on:

Authors > Keywords >

Abstract

Author keywords

Sustainable Development Goals 2023

SciVal Topics

Abstract

Real-time sensors in airports, power plants, intelligent manufacturing, and healthcare systems make multivariate time-series anomaly detection more critical. Two significant obstacles remain. First, data organisations isolate sensitive data on islands and train high-performance anomaly detection models to preserve privacy and security. Data organisations have statistical heterogeneity. A unified anomaly detection methodology fails with personalised data. Blochchain-aware federated anomaly detection framework (BcFad) for multivariate time series data. BcFad uses the federated learning architecture to aggregate data while respecting privacy and fine-tuning a reasonably personalised model. BcFad improves F1 scores by 6.9% relative to the baseline technique in NASA spacecraft dataset experiments. © 2024 selection and editorial matter, Arvind Dagur, Karan Singh, Pawan Singh Mehra & Dhirendra Kumar Shukla; individual chapters, the contributors.

SciVal Topics

Author keywords

Anomaly detection; Blockchain; Federated learning; Multivariant data

Sustainable Development Goals 2023 ⓘ New

SciVal Topics ⓘ

References (20)

View in search results format >

Blockchain-aware Federated Anomaly Detection scheme for Multivariate Data

Dr. V Selvakumar[1], Renato R. Maaliw III[2], Ravi Mohan Sharma[3], Rajvardhan Oak[4], Dr. Pavitar Parkash Singh[5], Dr. Ashok Kumar[6]

[1]Assistant Professor, Department of Maths and Statistics, Bhavan's Vivekananda College of Science, Humanities and Commerce, Hyderabad-94, Telangana

[2]College of Engineering, Southern Luzon State University, Lucban, Quezon, Philippines

[3]Associate professor , Department of Computer Science and Applications, Makhanalal Chaturvedi National University of Journalism and Communication, Bhopal, Madhya Pradesh, 462011

[4]Department of Computer Science, University of California Davis, USA

[5]Dean, Mittal School of Business, Lovely Professional University, Phagwara, Punjab

[6]Assistant Professor, Department of Computer Science, Banasthali Vidyapith, Rajasthan-304022, India

Mail: [1]vskselva79@gmail.com,[2]rmaaliw@slsu.edu.ph, [3]ravi@mcu.ac.in,

[4]rvoak@ucdavis.edu, [5]pavitar.19476@lpu.co.in, [6]kuashok@banasthali.in

Abstract: Real-time sensors in airports, power plants, intelligent manufacturing, and healthcare systems make multivariate time-series anomaly detection more critical. Two significant obstacles remain. First, data organisations isolate sensitive data on islands and train high-performance anomaly detection models to preserve privacy and security. Data organisations have statistical heterogeneity. A unified anomaly detection methodology fails with personalised data. Blochchain-aware federated anomaly detection framework (BcFad) for multivariate time series data. BcFad uses the federated learning architecture to aggregate data while respecting privacy and fine-tuning a reasonably personalised model. BcFad improves F1 scores by 6.9% relative to the baseline technique in NASA spacecraft dataset experiments.

Keywords: Anomaly detection, Blockchain, Federated learning,Multivariant data.

1. Introduction

Real-world applications of multivariate time series data have been widespread in numerous fields, such as weather data analysis and forecasting [1], healthcare [2], finance [3], etc. [4-5]. Anomaly detection is an important problem in multivariate time series analysis, the purpose is to detect sequence data that does not meet the expected behaviour, and it is one of the critical technologies of data mining. As deep learning has shown significant advantages in learning feature representations for complex data [6], anomaly detection using deep learning methods has received increasing attention in recent years. For instance, Deep Autoencoder Gaussian Mixture Model (DAGMM) [7-9] models the density distribution of multidimensional data using both Deep Autoencoder and Gaussian Mixture Model. In modelling temporal correlation in time series, LSTM has attained a high level of generalisation.

However, anomaly detection in multivariate time-series data still faces challenges. Taking flight data anomaly detection as an example, flight data is a specific multivariate time-series data; effective anomaly detection will improve the safety and reliability of aviation systems and improve maintenance action organization after landing. However, flight data is highly commercially confidential, causing data barriers between general airlines; at the same time, The probability distributions of flight data provided by aircraft of different sorts and tasks are vastly varied, Consequently, there is no universal detection model that can be used in all contexts.

This research introduces FedPAD, a personalised federated learning system for anomaly detection of multivariate time-series data, in response to the aforementioned issues. FedPAD can solve data silos and personalization problems at the same time. Through federated learning [10] and homomorphic encryption [11], FedPAD aggregates data from different institutions, builds a high-performance deep anomaly detection model in the cloud, and protects private data well. After the cloud model is established, FedPAD uses fine-tuning (fine-tuning) further to train a personalized anomaly detection model for each institution.

2. Related work

2.1 Prediction-based multivariate time-series anomaly detection model

LSTM-NDT [12] uses an extended short-term memory network to achieve high-performance time-series data prediction while ensuring the interpretability of the whole system. After the model generates forecast data, residuals are evaluated using a nonparametric, dynamic thresholding method. Specifically, let y_i be the signal value of the i-th time step of the input sequence and y_i' be the signal value of the i-th time step of the output sequence predicted by the model. Then, the prediction error is $e_i = y_i - y_i'$, with multiple time step error to compute the threshold sequence ε:

$$\varepsilon = \mu(e_s) + z\sigma(e_s) \quad (1)$$

Where e_s is the error sequence over multiple time steps, μ(·) is the mean, σ(·) is the standard deviation, and z is the weighting coefficient. The threshold ε_i of each time step is dynamically changed, depending on the maximum value of the previous entire threshold sequence ε; the calculation formula is as follows:

$$\varepsilon_i = argmax(\varepsilon) = \frac{\frac{\Delta\mu(e_s)}{\mu(e_s)}+\frac{\Delta\sigma(e_s)}{\sigma(e_s)}}{|e_a|+|E_{seq}|^2} \quad (2)$$

$$\Delta\mu(e_s) = \mu(e_s) - \mu(\{e_i \in e_s | e_i < \varepsilon_i\}) \quad (3)$$

$$\Delta\sigma(e_s) = \sigma(e_s) - \sigma(\{e_i \in e_s | e_i < \varepsilon_i\}) \quad (4)$$

Among them, e_a is the error value of the abnormal sequence, E_{seq} is the error value of the continuous abnormality in the abnormal sequence, and the expression is as follows:

$$e_a = \{e_i \in e_s | e_i > \varepsilon_i\} \quad (5)$$

$$E_{seq} = \{e_a\} \quad (6)$$

In addition, the pruning method is also used to reduce false positives, and the maximum value e_{max} in all error sequences is arranged in descending order to obtain e_s. Then the sequence is traversed to calculate the drop percentage d^i:

$$d^i = \frac{e_{max}^{i-1} - e_{max}^{i}}{e_{max}^{i-1}} \quad (7)$$

If di exceeds a minimum percentage p, the associated outlier series remain abnormal; if neither d^i nor all subsequent percentage declines exceed p, these erroneous series are reclassified as usual.

2.2 Federated Learning and Personalization

Google first proposed federated learning in 2016 [13]. It is built with the intention of facilitating secure data exchange among a large number of participants or computing nodes while also protecting individual privacy and adhering to all applicable laws and regulations. Many fields can benefit from federated learning due to its ability to efficiently address the issue of data islands; for instance, it has demonstrated strong performance and robustness when used to the next prediction problem on mobile devices [14]. A mountable system for massive federated mobile learning is proposed in the literature [15]. Using a federated learning framework that was edge-accelerated, the authors of [16] were able to attaincompetent recommendation concert while protecting users' anonymity in the POI recommendation task.

The primary motivation for institutions to participate in federated learning is to obtain better models. However, the global model acquired by federated learning may not be suitable for the detection requirements of institutions with sufficient local data to train efficient models. Actors may not be able to improve model performance via federated learning for many tasks, as it has been proven in the literature [17] that the globally shared model is not as accurate as the local model. Further, the global model's validity has been questioned in the published literature [18]. Non-IID client data makes it challenging to build a universal model that can be applied to all clients. To overcome this heterogeneity problem, we demonstrate that a standard federated learning approach, called Federated Averaging (FedAvg), can deal with certain types of non-IID data. In the face of severely skewed data distributions, however, FedAvg can lead to significant performance loss. Specifically, on one hand, FedAvg will produce a model with inferior performance compared to the centralised technique [19], as non-IID data will cause weight disparities between the federated learning process and the standard centralised training procedure. To contrast, FedAvg can only learn broad aspects from the data and is incapable of learning task-specific details.

To deal with the challenges of statistical heterogeneity and non-IID data, personalized global models become necessary. Most personalization techniques [20] usually consist of two steps. In the first step, an international model is trained collaboratively. In the second step, the global model is personalized for each client using the client's private data. FedPer was first proposed in the literature [20], wherein it was argued that the deep learning model should be viewed as the base + personalization layer, with the base layer functioning as a shared layer, and trained using existing federated learning techniques.

In contrast, the personalization layer is trained locally. By first training a global model via conventional federated learning and then sending it to all devices, as described in the aforementioned literature [19], it will be possible for each device to create a unique model by fine-tuning the global model with its own local data.Literature. [18] make a trade-off between traditional global and regional models, where each device can learn a local model from its local data without any communication. To achieve personalization, Literature. [20] did not calculate the average value of model parameters like FedAvg. Still, they figured out how much a client could benefit from another client's model aggregation and obtained the optimal value of each client—weighted Model Portfolio.

3. Method of this paper

3.1 Problem Definition

$\{S_1, S_2, \cdots, S_N\}$represent time series data from N different institutions $\{Q_1, Q_2, \cdots, Q_N\}$, and the data of different institutions have different distributions. Traditional centralized methods use global data S = $S_1 \cup S_2 \cup \ldots \cup S_N$ to train a unified model MALL. This work aims to train a federated anomaly detection model MFED using all available data. No company's data will be shared with another company's during the model training process. To achieve a detection rate that is on par with or better than that of MALL, it is necessary to train the federated anomaly detection model (MFED).

3.2 Framework Overview

FedPAD aims to achieve high-performance anomaly detection through personalized federated learning while protecting privacy. Figure 1 gives an overview of the framework, taking flight data anomaly detection as an example (it can be extended to other scenarios), assuming that there are three general aviation companies, each with different aircraft types. The framework mainly includes four processes. Initially, train the cloud model on the server using the publicly available dataset. The cloud model is subsequently disseminated to all institutions, with each institution introducing its own model utilising local data on top of the cloud model. Then, return the organization's model to the cloud and update the cloud's model using FedAvg. Repeat the preceding steps until the model converges or the specified training rounds are reached. Last but not least, each institution can utilise cloud models and local data to further train tailored models. Being that there is a significant discrepancy between the global statistics and the institution's local data, a method of fine-tuning is employed to better match the model to the local data. Through homomorphic encryption, all parameter-sharing functions will not leak user information throughout the procedure.

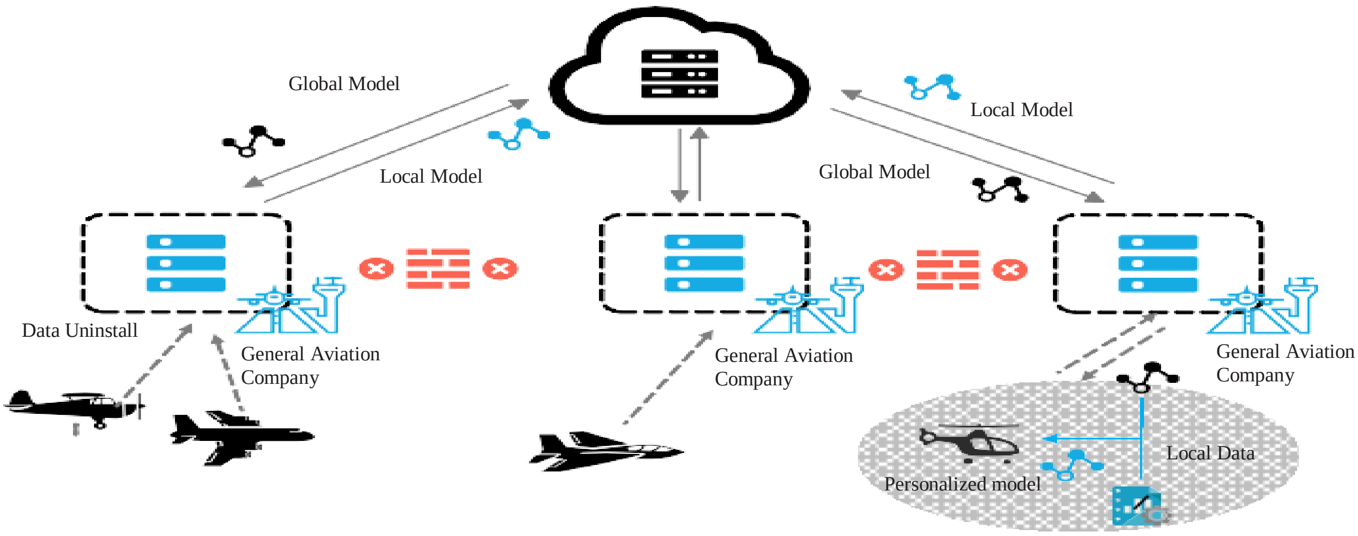


Figure 1 Overview of the FedPAD framework

3.3 Federated Learning

In order to solve the problem of data isolation and facilitate the training and distribution of distributed encryption models, FedPAD employs the federated learning paradigm. There are two main parts to this procedure: gathering models from the cloud and teaching models in an institution. FedPAD uses a neural network based on LSTM time series prediction as both the cloud and institutional model. When given access to institutional data, LSTM performs in-depth feature learning. The formulas (7) and (8) below illustrate the educational goals of both the cloud and campus-based models.

$$\arg\min_{w,b} L = \sum_{i=1}^{n} l(y_i, f_s(x_i)) \tag{7}$$

$$\arg\min_{w^j,b^j} L = \sum_{i=1}^{n^j} l(y_i^j, f_s(x_i^j))$$
(8)

Among these,ω and b stand for all of the factors that need to be learned, which include weights and biases, l(·, ·) represents the loss function, j represents the organization number, (x_i, y_i) and $(x_i^j, y_i^j\ )$ represent the Time-series data instances of j institutions, n and n j represent the size of the data set.

Once each user's model f j has been trained, it is uploaded to the cloud. Get the average model $f_S'$by aligning user models using the federated averaging approach [12] and averaging M user models during each training cycle.

$$f_S'(w,b) = \frac{1}{M}\sum_{m=1}^{M} f_{jm}(w,b) \tag{10}$$

3.4 Personalized Learning

The issue of data silos can be helped along by federated learning, which makes it possible for FedPAD to create anomaly detection models by making use of all institutional data. The statistical diversity of the data is an extra crucial factor that has an effect on performance. Due to the disparity in data distribution between a single institution and global data, the performance of directly employing the cloud model on a specific institution is still subpar.

The cloud-based generic model can only learn coarse features from all institutions and cannot learn fine-grained characteristics from the data of specific institutions. Because they emphasise learning standards and lower-level features, Yosinski et al. [13] proved that lower-level features in deep neural networks are highly transferrable. The network's higher layers will learn more specific elements. Consequently, after receiving the cloud model, the institution employs the approach of fine-tuning to actualize the personalised institution model, as depicted in Figure 2. The neural network consists of two LSTM layers, two Dropout layers, one Dense layer, and one Linear layer. We take in multivariate time series data and output forecasted time series data. While altering the parameters of the upper layers, FedPAD keeps the lowest layer settings (LSTM and Dropout) constant (Dense and Linear).

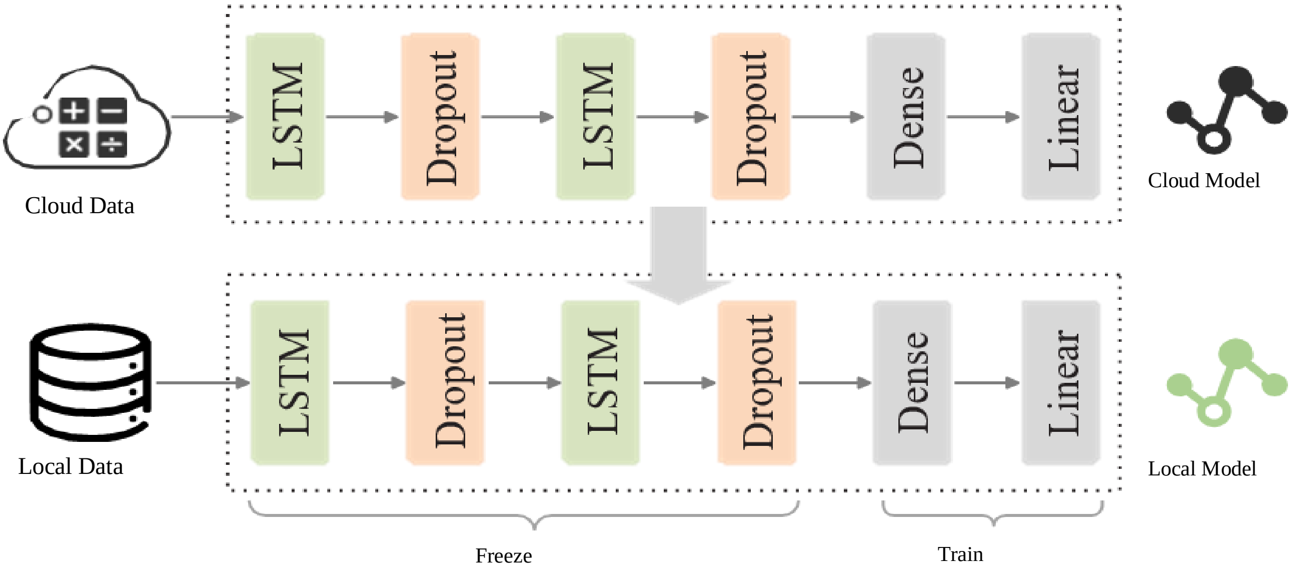


Figure 2 FedPAD fine-tuning process

3.5 Algorithm process

The model training process of FedPAD is introduced in Algorithm 1. When an institution generates new data, FedPAD can update the institution and cloud models. Therefore, the longer you use FedPAD, the better your model will perform.

```
Algorithm 1 FedPAD model training process
Input: public dataset D0, datasets $\{Q_1, Q_2, \cdots, Q_j\}$ from different institutions
Output: personalized models $\{ f_1, f_2, \cdots, f_j\}$ for different institutions
1. Use the public dataset to train the LSTM model $f_S$ in the cloud
2. FOR round =1,2, $\cdots r$ DO
```

3. Send the cloud model f_S to all institutions

4. Each institution uses local data D_j to train its model f_j and uploads the model parameters (ω, b) to the cloud

5. The cloud uses the federated average algorithm to aggregate all institutional model parameters and update the cloud model

6. END FOR

7. Each institution uses local data for fine-tuning to obtain a personalized model

4. Experiment and result analysis

4.1 Dataset

The description of the data set is shown in Table 1.

Table 1 FedPAD model training process

Data set	Number of channels	Feature dimension	Number of training sets	Number of test sets	Abnormal rate/%
SMAP	56	30	135179	427634	14.45
MSL	25	60	58328	73754	11.12

4.2 Evaluation Indicators

To make a direct comparison with the benchmark proposed by LSTM-NDT [11-12], the Point-based detection index commonly used in sequence data anomaly detection tasks is used; that is, when the predicted anomaly overlaps with the actual value, it is recorded as a true positive, and the expected anomaly and any true value When there is no intersection between the values, it is recorded as false positive. When there is no intersection between the real deal and any predicted value, it is registered as a false negative. Among them, the calculation of precision rate (Precision), recall rate (Recall) and F1 value are the same as the general detection task :

$$Precision = \frac{TP}{TP+FP}$$

(11)

$$Recall = \frac{TP}{TP+FN}$$

(12)

$$F1 = \frac{Precision \times Recall}{Precision + Recall} \quad (13)$$

Among them, TP is true positive, FP is false positive, TN is true negative, and FN is false negative.

4.3 Experimental setup

The time series data between channels in the NASA spacecraft data set have the same feature dimension. Still, there is a significant difference in feature distribution, that is, statistical heterogeneity, which is in line with the background of this paper. Therefore, each channel's data is regarded as an institution's data; that is, there are 55 institution nodes in the experiment on SMAP, and there are 27 institution nodes in the investigation on MSL for FedPAD model training. The LSTM-NDT method uses each channel data to train a model separately. The same model architecture and parameters as LSTM-NDT are used for a more intuitive comparison, as shown in Table 2.

Table 2 Model architecture and parameters

Parameter name	Value
Hidden layer	3
Hidden layer unit	82
Sequence length	255
Cycle	36

4.4 Result analysis

FedPAD is compared with LSTM-NDT, while the performance using only federated learning (FED) is recorded.

As shown in Table 3, compared with LSTM-NDT, only using federated learning for model training can solve the problem of data islands. Still, due to the statistical heterogeneity of data, the predictive performance of the FED model appears on both datasets. The overall forecast error increased by 1.2 percentage points. At the same time, the general prediction error of FedPAD has dropped by 1.6 percentage points, which is better than the LSTMNDT method. This is because federated learning can indirectly learn data characteristics from multiple institutions and then train a better model, and through fine-tuning, the model becomes better. Be more personalized and more adaptable to the data characteristics of each institution.

Table 3 Telemetry prediction error

Method	SMAP	MSL	Total
LSTM-NDT	5.7	7.4	6.5
FED	7.6	7.8	7.8
FedPAD	4.5	5.9	5.4

In the LSTM-NDT method, the pruning parameter p is vital to control the precision rate and the recall rate. The trade-off between the precision and recall rates is achieved by adjusting the parameter p. In the experiments in this paper, p is used as a control variable to compare the performance of the three methods in anomaly detection in two data sets. As shown in Figures 3 and 4, under different parameters p, the performance of the FED method is the most unstable, and the precision rate and recall rate cannot reach a high level simultaneously. On the other hand, the Fed-PAD method is slightly lower than LSTM-NDT when the value of p in the SMAP dataset is low. In other cases, the precision and recall are higher than LSTM-NDT, which benefits from the improvement of fine-tuning. Predictive performance of LSTM models.

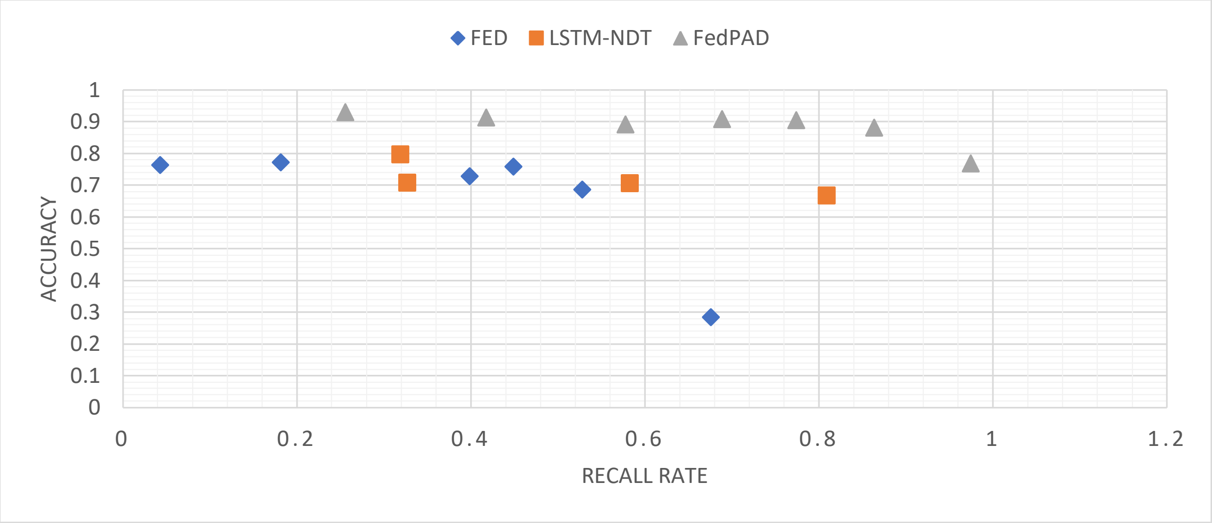


Figure 3 Performance comparison of the three methods in the MSL dataset

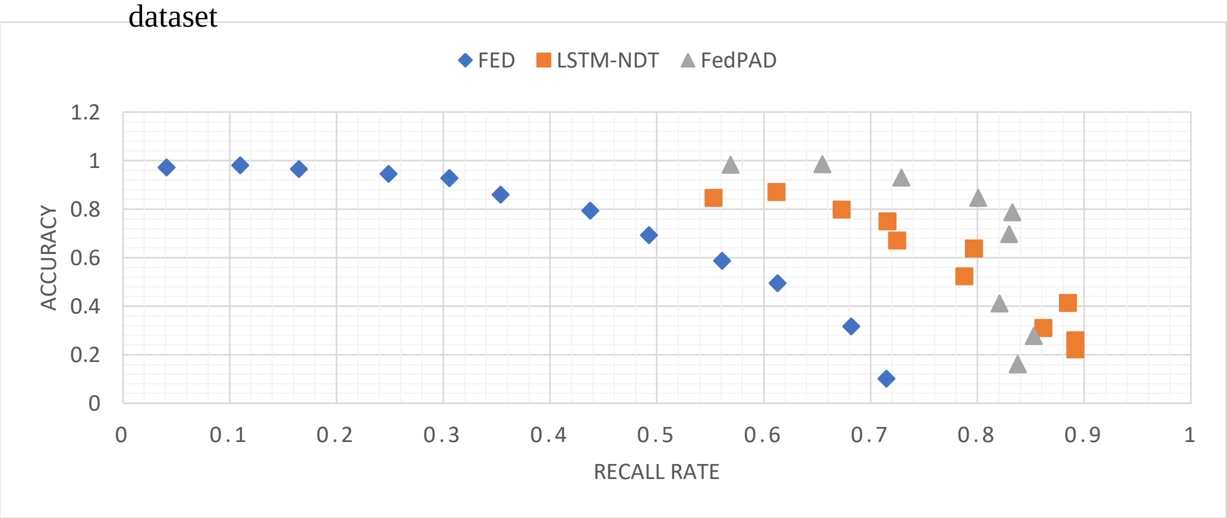


Figure 4 Performance comparison of the three methods in the SMAP dataset

Tables 4 and 5 show the average performance of the three methods under different parameters, and p is recorded. Compared with LSTM-NDT, FED's F1 scores on the two datasets decreased by 4.3% and 10.9%, respectively, and FedPAD's anomaly detection F1 scores increased by 10.1% on the MSL dataset and 3.6% on the SMAP dataset. the average F1 score increased by 6.9%. This again demonstrates the effectiveness of FedPAD in improving anomaly detection performance. The reason is that during the federated learning and fine-tuning process, the anomaly detection model on each data institution in FedPAD can learn the data characteristics of other institutions, which improves the reasoning performance of the model. Using a unified federated learning model is more prone to the problem of increased false positive rate or reduced model robustness. FedPAD can solve these problems by building a more personalized anomaly detection model for each data institution through fine-tuning.

Table 4 Average performance comparison on the MSL dataset

Method	Precision	Recall	F1

LSTM-NDT	0.88	0.56	0.62
FED	0.84	0.54	0.65
FedPAD	0.94	0.66	0.77

Table 5. Average performance comparison on the SMAP dataset

Method	Precision	Recall	F1
LSTM-NDT	0.87	0.86	0.72
FED	0.83	0.64	0.74
FedPAD	0.88	0.82	0.83

5. Conclusion

For anomaly detection of multivariate time series data, an anomaly detection framework FedPAD based on personalized federated learning is proposed. Based on the federated learning framework, FedPAD can learn the time series data characteristics of different institutions without disclosing data and privacy and use local data to model fine-tuning at the respective institutions to obtain personalized detection models. Experiments show that by detecting anomalies in NASA spacecraft data, FedPAD improves anomaly detection F1 scores by 6.9% compared to baseline methods.

Reference:

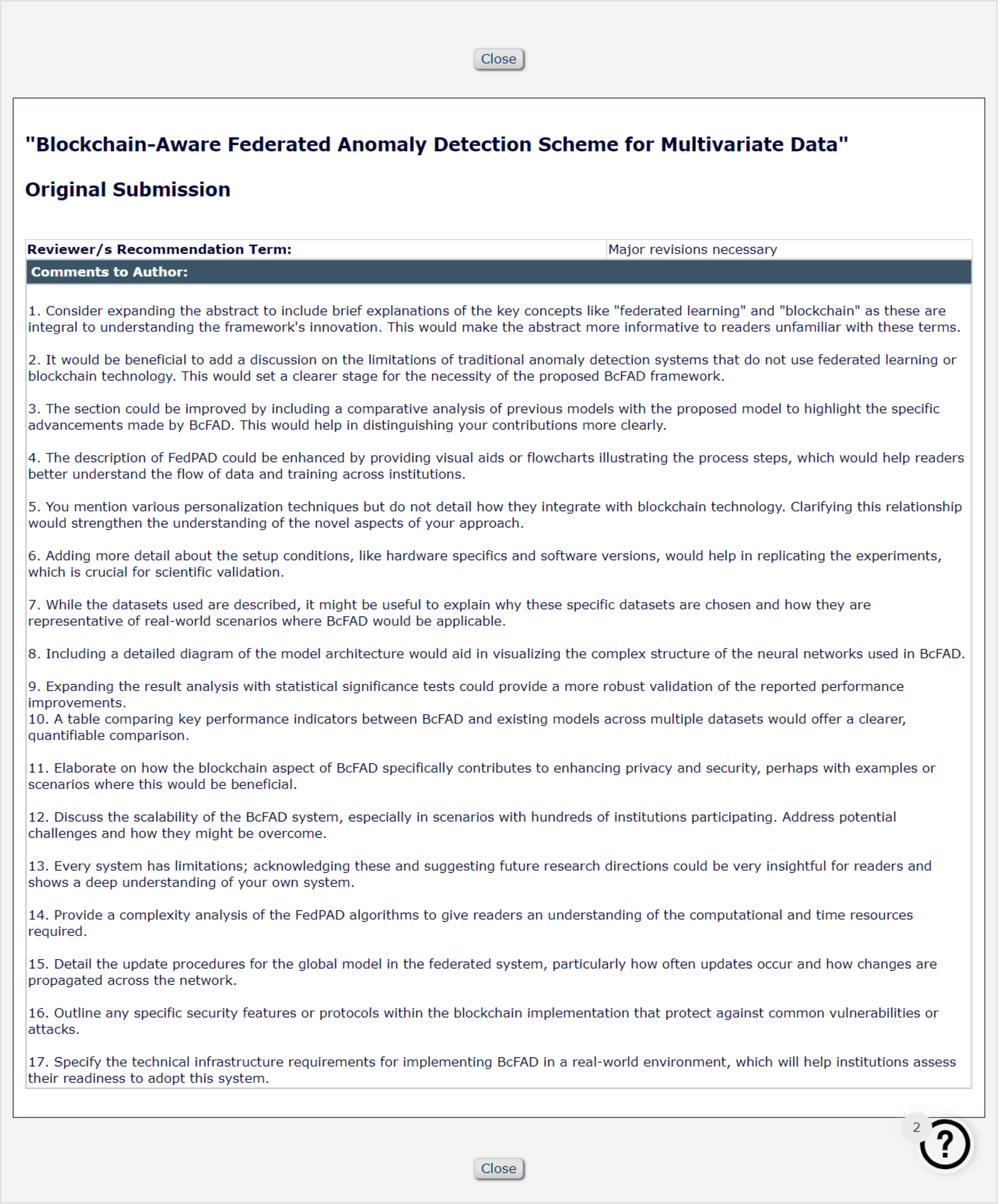
1. F. Tang, C. Wen, L. Luo, M. Zhao and N. Kato, "Blockchain-Based Trusted Traffic Offloading in Space-Air-Ground Integrated Networks (SAGIN): A Federated Reinforcement Learning Approach," in IEEE Journal on Selected Areas in Communications, vol. 40, no. 12, pp. 3501-3516, Dec. 2022, doi: 10.1109/JSAC.2022.3213317.
2. Y. Qu, S. R. Pokhrel, S. Garg, L. Gao and Y. Xiang, "A Blockchained Federated Learning Framework for Cognitive Computing in Industry 4.0 Networks," in IEEE Transactions on Industrial Informatics, vol. 17, no. 4, pp. 2964-2973, April 2021, doi: 10.1109/TII.2020.3007817.
3. D. C. Nguyen, S. Hosseinalipour, D. J. Love, P. N. Pathirana and C. G. Brinton, "Latency Optimization for Blockchain-Empowered Federated Learning in Multi-Server Edge Computing," in IEEE Journal on Selected Areas in Communications, vol. 40, no. 12, pp. 3373-3390, Dec. 2022, doi: 10.1109/JSAC.2022.3213344.
4. L. Cui, X. Su and Y. Zhou, "A Fast Blockchain-Based Federated Learning Framework With Compressed Communications," in IEEE Journal on Selected Areas in Communications, vol. 40, no. 12, pp. 3358-3372, Dec. 2022, doi: 10.1109/JSAC.2022.3213345.
5. Y. Miao, Z. Liu, H. Li, K. -K. R. Choo and R. H. Deng, "Privacy-Preserving Byzantine-Robust Federated Learning via Blockchain Systems," in IEEE Transactions on Information Forensics and Security, vol. 17, pp. 2848-2861, 2022, doi: 10.1109/TIFS.2022.3196274.
6. A. P. Kalapaaking, I. Khalil, M. S. Rahman, M. Atiquzzaman, X. Yi and M. Almashor, "Blockchain-Based Federated Learning With Secure Aggregation in Trusted Execution Environment for Internet-of-Things," in IEEE Transactions on

Industrial Informatics, vol. 19, no. 2, pp. 1703-1714, Feb. 2023, doi: 10.1109/TII.2022.3170348.

7. K. Toyoda, J. Zhao, A. N. S. Zhang and P. T. Mathiopoulos, "Blockchain-Enabled Federated Learning With Mechanism Design," in IEEE Access, vol. 8, pp. 219744-219756, 2020, doi: 10.1109/ACCESS.2020.3043037.
8. P. Wang et al., "Blockchain-Enhanced Federated Learning Market With Social Internet of Things," in IEEE Journal on Selected Areas in Communications, vol. 40, no. 12, pp. 3405-3421, Dec. 2022, doi: 10.1109/JSAC.2022.3213314.
9. N. Quang Hieu, T. A. Tran, C. L. Nguyen, D. Niyato, D. I. Kim and E. Elmroth, "Deep Reinforcement Learning for Resource Management in Blockchain-Enabled Federated Learning Network," in IEEE Networking Letters, vol. 4, no. 3, pp. 137-141, Sept. 2022, doi: 10.1109/LNET.2022.3173971.
10. R. Kumar et al., "Blockchain-Federated-Learning and Deep Learning Models for COVID-19 Detection Using CT Imaging," in IEEE Sensors Journal, vol. 21, no. 14, pp. 16301-16314, 15 July15, 2021, doi: 10.1109/JSEN.2021.3076767.
11. M. Abdel-Basset, N. Moustafa, H. Hawash, I. Razzak, K. M. Sallam and O. M. Elkomy, "Federated Intrusion Detection in Blockchain-Based Smart Transportation Systems," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 3, pp. 2523-2537, March 2022, doi: 10.1109/TITS.2021.3119968.
12. V. Mothukuri, R. M. Parizi, S. Pouriyeh, A. Dehghantanha and K. -K. R. Choo, "FabricFL: Blockchain-in-the-Loop Federated Learning for Trusted Decentralized Systems," in IEEE Systems Journal, vol. 16, no. 3, pp. 3711-3722, Sept. 2022, doi: 10.1109/JSYST.2021.3124513.
13. Z. Peng et al., "VFChain: Enabling Verifiable and Auditable Federated Learning via Blockchain Systems," in IEEE Transactions on Network Science and Engineering, vol. 9, no. 1, pp. 173-186, 1 Jan.-Feb. 2022, doi: 10.1109/TNSE.2021.3050781.
14. J. Sun, Y. Wu, S. Wang, Y. Fu and X. Chang, "Permissioned Blockchain Frame for Secure Federated Learning," in IEEE Communications Letters, vol. 26, no. 1, pp. 13-17, Jan. 2022, doi: 10.1109/LCOMM.2021.3121297.
15. J. Qi, F. Lin, Z. Chen, C. Tang, R. Jia and M. Li, "High-Quality Model Aggregation for Blockchain-Based Federated Learning via Reputation-Motivated Task Participation," in IEEE Internet of Things Journal, vol. 9, no. 19, pp. 18378-18391, 1 Oct.1, 2022, doi: 10.1109/JIOT.2022.3160425.
16. Y. Wang, H. Peng, Z. Su, T. H. Luan, A. Benslimane and Y. Wu, "A Platform-Free Proof of Federated Learning Consensus Mechanism for Sustainable Blockchains," in IEEE Journal on Selected Areas in Communications, vol. 40, no. 12, pp. 3305-3324, Dec. 2022, doi: 10.1109/JSAC.2022.3213347.
17. X. Cheng, W. Tian, F. Shi, M. Zhao, S. Chen and H. Wang, "A Blockchain-Empowered Cluster-Based Federated Learning Model for Blade Icing Estimation on IoT-Enabled Wind Turbine," in IEEE Transactions on Industrial Informatics, vol. 18, no. 12, pp. 9184-9195, Dec. 2022, doi: 10.1109/TII.2022.3159684.
18. W. Zhang et al., "Blockchain-Based Federated Learning for Device Failure Detection in Industrial IoT," in IEEE Internet of Things Journal, vol. 8, no. 7, pp. 5926-5937, 1 April1, 2021, doi: 10.1109/JIOT.2020.3032544.
19. Y. Lu, X. Huang, K. Zhang, S. Maharjan and Y. Zhang, "Low-Latency Federated Learning and Blockchain for Edge Association in Digital Twin Empowered 6G

Networks," in IEEE Transactions on Industrial Informatics, vol. 17, no. 7, pp. 5098-5107, July 2021, doi: 10.1109/TII.2020.3017668.

20. F. Ayaz, Z. Sheng, D. Tian and Y. L. Guan, "A Blockchain Based Federated Learning for Message Dissemination in Vehicular Networks," in IEEE Transactions on Vehicular Technology, vol. 71, no. 2, pp. 1927-1940, Feb. 2022, doi: 10.1109/TVT.2021.3132226.



"Blockchain-Aware Federated Anomaly Detection Scheme for Multivariate Data"

Original Submission

Reviewer/s Recommendation Term:	Major revisions necessary

Comments to Author:

1. Consider expanding the abstract to include brief explanations of the key concepts like "federated learning" and "blockchain" as these are integral to understanding the framework's innovation. This would make the abstract more informative to readers unfamiliar with these terms.

2. It would be beneficial to add a discussion on the limitations of traditional anomaly detection systems that do not use federated learning or blockchain technology. This would set a clearer stage for the necessity of the proposed BcFAD framework.

3. The section could be improved by including a comparative analysis of previous models with the proposed model to highlight the specific advancements made by BcFAD. This would help in distinguishing your contributions more clearly.

4. The description of FedPAD could be enhanced by providing visual aids or flowcharts illustrating the process steps, which would help readers better understand the flow of data and training across institutions.

5. You mention various personalization techniques but do not detail how they integrate with blockchain technology. Clarifying this relationship would strengthen the understanding of the novel aspects of your approach.

6. Adding more detail about the setup conditions, like hardware specifics and software versions, would help in replicating the experiments, which is crucial for scientific validation.

7. While the datasets used are described, it might be useful to explain why these specific datasets are chosen and how they are representative of real-world scenarios where BcFAD would be applicable.

8. Including a detailed diagram of the model architecture would aid in visualizing the complex structure of the neural networks used in BcFAD.

9. Expanding the result analysis with statistical significance tests could provide a more robust validation of the reported performance improvements.
10. A table comparing key performance indicators between BcFAD and existing models across multiple datasets would offer a clearer, quantifiable comparison.

11. Elaborate on how the blockchain aspect of BcFAD specifically contributes to enhancing privacy and security, perhaps with examples or scenarios where this would be beneficial.

12. Discuss the scalability of the BcFAD system, especially in scenarios with hundreds of institutions participating. Address potential challenges and how they might be overcome.

13. Every system has limitations; acknowledging these and suggesting future research directions could be very insightful for readers and shows a deep understanding of your own system.

14. Provide a complexity analysis of the FedPAD algorithms to give readers an understanding of the computational and time resources required.

15. Detail the update procedures for the global model in the federated system, particularly how often updates occur and how changes are propagated across the network.

16. Outline any specific security features or protocols within the blockchain implementation that protect against common vulnerabilities or attacks.

17. Specify the technical infrastructure requirements for implementing BcFAD in a real-world environment, which will help institutions assess their readiness to adopt this system.

FACULTY POSITION RECLASSIFICATION FOR SUCS

(DBM-CHED Joint Circular No. 3, series of 2022)

CERTIFICATION OF PERCENTAGE CONTRIBUTION

(Research Output with Multiple Authors)

Title of Research: Blockchain-Aware Federated Anomaly Detection Scheme for Multivariate Data
Type of Research Output: Book Chapter (Scopus-Indexed, CRC Press (Taylor & Francis)

Instruction: Supply ALL the names of the authors involved in the publication of Research output and indicate the contribution of each author in percentage. Each author shall sign the conforme column if he/she agrees with the distribution. The conforme should be signed by all the authors in order to be considered. Please prepare separate Certification for each output.

	Name of Authors	Current Affiliations	% Contribution	Conforme *(Sign if you agree with the % distribution)*
1	V. Selvakumar	Vivekanada College of Science	16.67%	Selvakumar
2	Renato R. Maaliw III	SLSU	16.67%	RM3
3	Ravi Mohan Sharma	Chaturvedi National University	16.67%	
4	Rajvardhan Oak	University of California Davis	16.67%	Rajvardhan
5	Pavitar Parkash Singh	Lovely Professional University	16.67%	
6	Ashok Kumar	BV University	16.67%	
	** Should have a total of 100%*		100.00%	

Prepared by:

RM3

Renato R. Maaliw III, DIT
(Name and Signature)
Faculty

Certified by:

Nicanor L. Guinto, Ph.D
(Name and Signature)
Director, Office of Research Services