Trustworthy and Explainable Federated System for Extracting Descriptive Rules in a Data Streaming Environment

María Asunción Padilla-Rascón
Department of Computer Science
University of Jaén
E-23071 Jaén (Spain)
Andalusian Research Institute in Data Science and Computational Intelligence
University of Jaén
E-23071 Jaén (Spain)
mprascon@ujaen.es

Ángel Miguel García-Vico
Department of Computer Science
University of Jaén
E-23071 Jaén (Spain)
Andalusian Research Institute in Data Science and Computational Intelligence
University of Jaén
E-23071 Jaén (Spain)
agvico@ujaen.es

Cristóbal J. Carmona
Department of Computer Science
University of Jaén
E-23071 Jaén (Spain)
Andalusian Research Institute in Data Science and Computational Intelligence
University of Jaén
E-23071 Jaén (Spain)
Leicester School of Pharmacy
De Montfort University
LE1 7RH, Leicester, United Kingdom
ccarmona@ujaen.es

Abstract

In the information age, continuous streams of data from connected devices require intelligent models that ensure security, privacy and transparency. Federated learning enables knowledge sharing while adhering to the principles of trustworthy AI. This work synthesizes the *Trustworthy and Explainable Federated System for Extracting Descriptive Rules in a Data Streaming Environment (TEFeS-SDR)* [1] study, which introduces an evolutionary single-objective federated system for extracting descriptive rules while prioritizing privacy and security through advanced encryption techniques (binary, symmetric, and asymmetric). It ensures traceability and transparency, and experimental results confirm its resilience to concept drift while maintaining high quality models, advancing responsible AI by integrating explainability, security and efficiency.

XVI XVI Congreso Español de Metaheurísticas, Algoritmos Evolutivos y Bioinspirados (maeb 2025).

1 Introduction

The accelerated increase in continuous data generation from interconnected devices requires real-time learning models that preserve user privacy. Federated learning addresses this challenge by enabling distributed model training without sharing raw data, but it still faces issues related to security, transparency, and explainability.

This study presents a condensed overview of *Trustworthy and Explainable Federated System for Extracting Descriptive Rules in a Data Streaming Environment (TEFeS-SDR)* [1], a federated system designed to enhance security, explainability, and reliability in federated learning. It integrates an evolutionary algorithm based on emerging pattern mining (EPM) to generate interpretable rules while employing a hybrid encryption scheme (binary, symmetric, and asymmetric) for secure knowledge sharing. By combining the inherent explainability of EPM algorithms (absent in existing secure federated learning approaches [2, 3, 4]) with advanced security techniques not previously applied to federated learning with EPM[5, 6], TEFeS-SDR effectively balances explainability and security.

2 Methodology

The proposed algorithm TEFeS-SDR [1] is a hierarchical, federated rule-based model designed to extract explainable and trustworthy knowledge from dynamic systems at varying levels of granularity. Each local node processes its data stream using a single-objective evolutionary rule-based algorithm, generating local knowledge that is sent to a central fusion node, where it is consolidated into a global model. This model is subsequently shared back with local nodes, enabling mutual refinement.

To ensure privacy, raw data remains on each node, and only knowledge is shared. However, to mitigate potential risks, the system incorporates encryption mechanisms, including a keyring-based trust system for secure peer-to-peer exchange to restrict access to authorized recipients and binary encryption to prevent access to shared knowledge.

The system follows a hierarchical client-server architecture with two main components:

1. **Clients (local nodes)**: Low-power devices, such as Raspberry Pi, running a single-objective evolutionary fuzzy algorithm based on emerging pattern mining (EPM). This algorithm represents each individual as a pattern [7] encoded using triangular fuzzy linguistic labels (LLs) [8], applying genetic operators such as binary tournament selection [9], two-point crossover [10], and biased mutation specific for EPM algorithms [11]. Additionally, an elitist replacement scheme is included to retain the best solutions, along with a concept drift detection mechanism based on population quality.

2. **Server (fusion node)**: Receives and aggregates local knowledge into a global model using rule fusion techniques such as confidence filtering and token competition filter.

3 Experimental study

The experimental study conducted in our work [1] involves a central fusion node using token competition or confidence filtering, along with multiple Raspberry Pi devices simulating IoT or wearable nodes. The setup includes four Raspberry Pi 4 Model B devices as clients and a server running Ubuntu 23.04 with an Intel Core i7 processor. The main hypothesis is that the global model remains homogeneous despite concept drifts in the data stream.

3.1 Datasets

The study utilizes artificial data streams generated with MOA [12], consisting of 200 blocks of 5,000 instances, totalling one million instances per client, while the fusion node contains a validation dataset of 500,000 instances. Concept drifts, occur randomly between the 10th and 50th block of the current batch. Each stream is generated with a unique seed per device. Four data streams are used: Aggrawal, Mixed, RandomTree, and SEA, each with two classes and 9, 4, 10, and 3 attributes, respectively.

3.2 Parameters of the algorithm

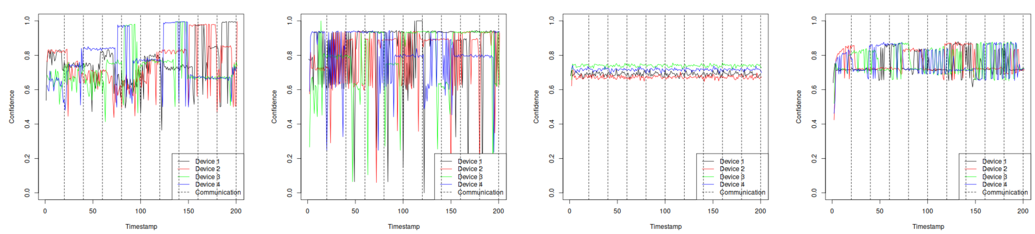
The parameters used are 3 fuzzy linguistic labels [13], a population size of 50, crossover and mutation probabilities of 0.7 and 0.05, and a maximum of 5,000 evaluations. The optimized objectives, using a weighted sum approach, are WRAccN, Support Difference, and Confidence. To detect concept drift, confidence (0.6) and TPR (0.1) thresholds are defined.

3.3 Results and analysis

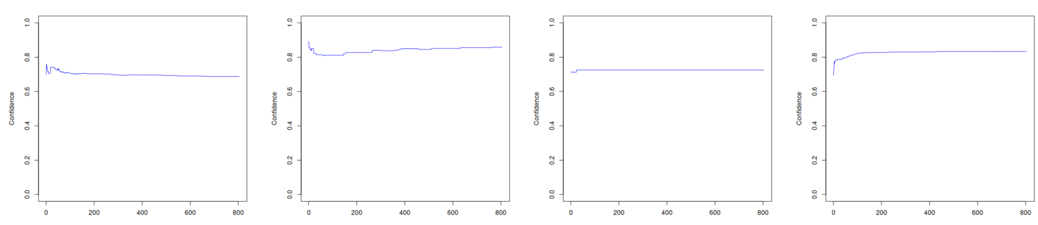
The evaluation of the final global model (Table 1) shows that the token competition fusion method reduces the number of variables compared to confidence-based filtering while maintaining a stable number of rules. This increases explainability without increasing complexity and good interpretability. Additionally, results from the confidence fusion method on datasets with concept drift (Figure 1 indicate that model confidence remains between 0.6 and 0.9, demonstrating strong performance. The global model remains stable despite concept drifts, confirming the hypothesis and highlighting the algorithm's robustness to data changes.

Table 1: Average results of the different fusion methods analysed

Timestamp	NumRules	NumVars	CONF	WRAcc	GR	FPR	TPR
Confidence	101.5	81.4995	0.6190	0.6195	0.5068	0.2682	0.8542
TokenCompetition	101.5	8.0788	0.6071	0.6077	0.4915	0.2684	0.8397



a b c d



a b c d

Figure 1: Comparison of the average confidence using confidence fusion method applied to a: Agrawal, b: Mixed, c: Random Tree, d: SEA datasets. The first row presents the results of the local models. The second row corresponds to the global model.

4 Conclusions

In an interconnected world generating vast data volumes, our work TEFeS-SDR [1] addresses security, reliability, and explainability challenges by integrating federated learning with an evolutionary algorithm based on emerging pattern mining, and a hybrid encryption scheme. This approach enables local data processing, enhancing privacy and minimizing transmission risks.

TEFeS-SDR also ensures explainability by extracting interpretable rules at local and global levels, fostering trust through traceable decision-making. Its encrypted and auditable knowledge transactions ensure transparency and regulatory compliance. Experimental results confirm its robustness against abrupt data changes, making it a reliable solution for dynamic environments.

References

[1] María Asunción Padilla Rascón, Ángel Miguel García-Vico, and Cristóbal J. Carmona. Trustworthy and explainable federated system for extracting descriptive rules in a data streaming environment. *Results in Engineering*, 25:104137, 2025.

[2] Othmane Marfoq, Giovanni Neglia, Laetitia Kameni, and Richard Vidal. Federated learning for data streams. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206, pages 8889–8924, 2023.

[3] Jaehong Yoon, Wonyong Jeong, Giwoong Lee, Eunho Yang, and Sung Ju Hwang. Federated continual learning with weighted inter-client transfer. In *International Conference on Machine Learning*, pages 12073–12086, 2021.

[4] Olusola Odeyomi and Gergely Zaruba. Differentially-private federated learning with long-term constraints using online mirror descent. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 1308–1313. IEEE, 2021.

[5] Pedro Miguel Sánchez Sánchez, Alberto Huertas Celdrán, Ning Xie, Gérôme Bovet, Gregorio Martínez Pérez, and Burkhard Stiller. Federatedtrust: A solution for trustworthy federated learning. *Future Generation Computer Systems*, 152:83–98, 2024.

[6] Jing Ma, Si-Ahmed Naas, Stephan Sigg, and Xixiang Lyu. Privacy-preserving federated learning based on multi-key homomorphic encryption. *International Journal of Intelligent Systems*, 37(9):5880–5901, 2022.

[7] Oscar Cordón, F Herrera, F Hoffmann, and L Magdalena. Evolutionary tuning and learning of fuzzy knowledge bases. *Genetic fuzzy systems*, 19, 2001.

[8] Lotfi Asker Zadeh. The concept of a linguistic variable and its application to approximate reasoning—i. *Information sciences*, 8(3):199–249, 1975.

[9] Brad L Miller, David E Goldberg, et al. Genetic algorithms, tournament selection, and the effects of noise. *Complex systems*, 9(3):193–212, 1995.

[10] John H Holland. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.

[11] Ángel Miguel García-Vico, Cristóbal José Carmona, Pedro González, and María José del Jesus. Moea-efep: Multi-objective evolutionary algorithm for extracting fuzzy emerging patterns. *IEEE Transactions on Fuzzy Systems*, 26(5):2861–2872, 2018.

[12] Albert Bifet, Geoff Holmes, Bernhard Pfahringer, Philipp Kranen, Hardy Kremer, Timm Jansen, and Thomas Seidl. Moa: Massive online analysis, a framework for stream classification and clustering. In *Proceedings of the First Workshop on Applications of Pattern Analysis*, volume 11 of *Proceedings of Machine Learning Research*, pages 44–50, 2010.

[13] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning—i. *Information Sciences*, 8(3):199–249, 1975.