Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Towards Transparent Rare Disease Monitoring: Federated Rule Extraction for Online Streaming NPC1 Data

María Asunción Padilla-Rascón

mprascon@ujaen.es

University of Jaén

Ángel Miguel García-Vico

University of Jaén

Martin Grootveld

De Montfort University

David Elizondo

De Montfort University

Cristóbal J. Carmona

University of Jaén

Research Article

Keywords: federated learning, evolutionary fuzzy systems, online learning, NPC1, rare disease detection, metabolomics, trustworthy AI

Posted Date: May 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9691952/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Towards Transparent Rare Disease Monitoring: Federated Rule Extraction for Online Streaming NPC1 Data

María Asunción Padilla-Rascón[1,2*], Ángel Miguel García-Vico[1,2], Martin Grootveld[3], David Elizondo[4], Cristóbal J. Carmona[1,2,3]

[1]Department of Computer Science, University of Jaén, Jaén, Andalucía, Spain.
[2]Andalusian Research Institute in Data Science and Computational Intelligence, Jaén, Andalucía, Spain.
[4]School of Computer Science and Informatics, Faculty of Technology, De Montfort University, The Gateway, Leicester, LE1 9BH, United Kingdom.
[3]Leicester School of Pharmacy, Faculty of Health and Life Sciences, De Montfort University, The Gateway, Leicester, LE1 9BH, United Kingdom.

*Corresponding author(s). E-mail(s): mprascon@ujaen.es;
Contributing authors: agvico@ujaen.es; mgrootveld@dmu.ac.uk; elizondo@dmu.ac.uk; ccarmona@ujaen.es;

Abstract

The early and accurate diagnosis of Niemann-Pick disease type C1 (NPC1) poses a fundamental challenge for machine learning: clinical datasets are exceptionally small, severely imbalanced, geographically fragmented, and subject to strict privacy regulations. Furthermore, existing black-box models are ill-suited to this regime, lacking the data efficiency and interpretability necessary for clinical validation. This paper presents OFEFS (Online Federated Evolutionary Fuzzy System), a novel framework that addresses these constraints by integrating federated learning, online incremental adaptation, and evolutionary fuzzy rule induction. OFEFS ensures data privacy by keeping patient records localized, while continuously updating a globally shared model without requiring full retraining. To handle the adverse data characteristics typical of rare diseases, the pipeline incorporates fuzzy-based feature selection and synthetic oversampling to

navigate extreme dimensionality and class imbalance. Experimental validation on a clinical cohort of NMR urine spectra demonstrates that the global federated model achieves a robust equilibrium between diagnostic sensitivity and clinical reliability, autonomously recovering rules that align with established biochemical pathways. Crucially, OFEFS synthesizes distributed knowledge into compact, human-readable linguistic rules (IF-THEN), providing clinicians with a transparent, trustworthy AI tool that preserves patient privacy and adapts dynamically to new clinical data streams.

Keywords: federated learning, evolutionary fuzzy systems, online learning, NPC1, rare disease detection, metabolomics, trustworthy AI

1 Introduction

Niemann–Pick disease type C1 (NPC1, OMIM 257220) is a rare neurodegenerative lysosomal storage disorder caused by autosomal recessive mutations in the NPC1 gene, accounting for approximately 95% of clinically diagnosed cases [1]. With an estimated incidence of 1 in 100,000–120,000 live births [2], the disease is characterised by abnormal lysosomal accumulation of unesterified cholesterol and sphingolipids, which triggers a cascade of neuroinflammation, neuronal apoptosis, and oxidative stress [3]. Diagnosis currently relies on a combination of clinical assessment, filipin staining, and genetic testing [4], while reliable early-stage biomarkers remain limited.

In this context, high-resolution proton (^{1}H) NMR spectroscopy of urine has emerged as a powerful non-invasive approach for biomarker discovery in NPC1. Previous studies have consistently reported alterations in key metabolic pathways, including tryptophan–nicotinamide metabolism, the kynurenine pathway, histidine metabolism, and branched-chain amino acids [5].

The kynurenine pathway is of particular pathophysiological relevance in this context: as the principal hepatic route for tryptophan catabolism, its chronic activation in NPC1 shifts metabolite flux away from neuroprotective compounds, such as kynurenate and xanthurenate, towards neurotoxic downstream products including quinolinate and 3-hydroxykynurenine, promoting excitotoxicity andneuroinflammation [6, 7]. Analogous disruptions have been reported in other lysosomal storage disorders such as Fabry and Gaucher's disease [8, 9], and elevated kynurenine-to-tryptophan ratios are now established as a systemic neuroinflammatory indicator in neurodegenerative conditions more broadly [10, 11]. Urinary NMR profiles in NPC1 further reflect disruptions in histidine metabolism and glycine conjugation pathways, converging on a coherent picture of systemic metabolic dysregulation amenable to non-invasive biomarker discovery [12, 13].

Despite these advances, the computational analysis of NPC1 metabolomics data presents several inherent challenges. First, the extreme rarity of the disease results in very limited sample sizes, typically characterised by a high-dimensional feature space with far fewer observations. Second, severe class imbalance, where heterozygous carrier controls substantially outnumber NPC1 patients, biases standard machine learning

models against the clinically relevant minority class, thereby reducing diagnostic sensitivity.

Beyond these data-related challenges, additional constraints arise from real-world clinical settings. Medical data for rare diseases is typically distributed across multiple institutions, each holding small, privacy-sensitive datasets. The centralised aggregation of such data is often infeasible due to regulatory frameworks such as the General Data Protection Regulation [14], which impose strict limitations on data sharing. Moreover, clinical data acquisition is inherently dynamic, as new patient samples become available over time. This requires learning systems capable of adapting incrementally without the need for complete retraining. Federated learning (FL) [15] addresses the privacy constraint by enabling collaborative model training without sharing raw data, while online learning [16] provides mechanisms for continuous adaptation to streaming data. The combination of both paradigms, known as online federated learning [17], therefore represents a promising and practically relevant framework for rare disease monitoring, although its effectiveness depends on thespecific characteristics of the underlying data and deployment constraints.

In addition to architectural considerations, the choice of learning paradigm is crucial in ultra-rare disease scenarios. Contemporary deep learning (DL) and other complex black-box models are typically data-hungry, although recent advances in transfer learning and regularization have partially mitigated these limitations in certain domains. Nevertheless, in ultra-rare disease scenarios such as NPC1, the extreme scarcity and imbalance of data remain significant challenges for such approaches. In the context of NPC1, where instances are profoundly scarce and heavily imbalanced, such approaches struggle to extract reliable patterns and are highly susceptible to overfitting. Furthermore, the European Commission's Ethics Guidelines for Trustworthy AI [18] establish that clinical AI systems must be transparent and human-interpretable, properties that black-box models fundamentally lack. Evolutionary fuzzy systems (EFS) [19] present a viable alternative in low-data settings, particularly when interpretability is required, although they may face limitations in scalability and stability depending on configuration and data characteristics. Not only do they demonstrate remarkable robustness in achieving promising predictive performance despite extreme data scarcity and class imbalance, but they are also inherently aligned with trustworthiness principles. They produce IF-THEN rules expressed in natural linguistic terms (e.g., "IF nicotinate is low AND trigonelline is low THEN NPC1"), which clinical practitioners can directly read, validate, and act upon in a transparent manner [20].

Although FL and EFS have been extensively studied in isolation, their integration, particularly in an online setting, remains largely unexplored. Specifically, there is a lack of frameworks capable of performing distributed, privacy-preserving learning while incrementally updating interpretable fuzzy rule-based models.

To address this gap, an online federated evolutionary fuzzy system (OFEFS) is proposed for the urinary metabolomic discrimination of NPC1. The proposed framework integrates: (i) data augmentation for robust dataset initialisation; (ii) a novel local evolutionary fuzzy algorithm based on a test-then-train paradigm; and (iii) a centralised online federated architecture capable of incremental adaptation via sliding windows. The principal contributions of this paper are:

- **The OFEFS Framework:** A novel architectural synthesis that synchronises online streaming learning with federated data processing and evolutionary fuzzy systems, specifically designed for non-stationary environments.
- **Privacy & Continuous Adaptation:** A centralized topology that ensures patient data confidentiality while adapting incrementally to new clinical data streams through a sliding-window approach, eliminating the need for computationally expensive global retraining.
- **Trustworthy AI & Interpretability:** The extraction of highly accurate, human-readable linguistic IF–THEN rules that provide clinicians with transparent, auditable insights, directly addressing the ethical requirements for trustworthy AI in medicine.
- **Robust Handling of Adverse Data:** A specialized pipeline incorporating Fuzzy Rough Feature Selection (FRFS) and NBalance oversampling to effectively tackle the extreme high-dimensionality ($p \gg n$) and severe class imbalance inherent to ultra-rare disease analytics.

The remainder of the paper is organized as follows. Section 2 provides a comprehensive review of the state of the art and related literature. Section 3 presents the core methodological contribution of this work, describing in detail the proposed OFEFS algorithm and its federated architecture. Section 4 is dedicated to the experimental study, covering the dataset characteristics, the applied preprocessing, and a detailed discussion of the results. Finally, Section 5 outlines the concluding remarks of the study.

2 State of the art and related works

Federated Learning (FL), originally proposed by McMahan et al. [15], enables the collaborative training of a global machine learning model across distributed clients without the need to share raw data. In this decentralized architecture, each participating node computes local model updates using its private dataset and transmits only gradients or model parameters to a central server, which subsequently aggregates them to refine the global model. This paradigm directly addresses the historical tension between data utility and privacy, which is particularly acute in the biomedical domain where patient data is inherently sensitive and subject to strict regulatory frameworks, such as the Health Insurance Portability and Accountability Act (HIPAA) in the US and the General Data Protection Regulation (GDPR) in Europe [14]. Consequently, comprehensive taxonomies by Yang et al. [21] and Rauniyar et al. [22] illustrate the successful deployment of FL across diverse medical applications. Highlighting the maturity of this direction, Zhang et al. [23] recently developed a fairness-aware and privacy-preserving FL framework designed to address resource heterogeneity within collaborative AI ecosystems.

Despite these advancements, a non-negligible gap remains concerning the application of FL to structured, tabular clinical data, the ubiquitous format produced by routine laboratory analyses, metabolomic profiling, and phenotypic registries. Tabular data present distinct algorithmic challenges compared to medical imaging: feature spaces are highly heterogeneous, sample sizes are typically small (particularly in rare

diseases), and data distributions across institutions often diverge substantially due to variations in measurement protocols and localized disease prevalence. Casella et al. [24] addressed this gap with MERGE, a multi-input architecture demonstrating that including structured features improves federated model generalizability. However, the challenge of non-independent and identically distributed (non-IID) tabular data becomes particularly severe in multi-institutional settings. This statistical heterogeneity causes local updates to diverge from the global optimum, degrading convergence [25]. Moreno-Barea et al. [26] illustrated this problem concretely within the context of Niemann-Pick type C1 (NPC1) disease, demonstrating that metabolomic datasets for such patients are inherently limited in sample size and severely class-imbalanced, conditions that heavily compound the difficulties of federated aggregation and demand highly resilient modeling approaches.

Beyond the statistical challenges of tabular data, the adoption of AI in clinical decision support has prompted a paradigm shift towards trustworthy AI, encompassing transparency, explainability, fairness, and robustness [27]. In a federated context, ensuring trustworthiness is inherently complicated: privacy guarantees and model transparency must be achieved without centralized data access. Chaddad et al. [28] underscore that explainability is a critical requirement for FL models applied to structured medical data, arguing that clinical interpretability holds equal importance to predictive accuracy. While explainability in FL is often approached through post-hoc techniques (e.g., SHAP), intrinsically interpretable models are increasingly preferred. Ducange et al. [29] demonstrated this by comparing federated fuzzy rule-based systems against neural networks equipped with SHAP in a Parkinson's case study, finding that fuzzy systems achieved competitive accuracy while offering superior, strictly compliant transparency.

Fuzzy Rule-Based Systems (FRBS) constitute one of the most established paradigms for interpretable machine learning. As argued by Rudin [30], interpretable-by-design models are vastly preferable to post-hoc explanations in high-stakes contexts, as they provide globally consistent explanations. Building upon this, EFS augment classical FRBS with evolutionary algorithms to simultaneously optimize the structure, parameters, and number of rules, balancing predictive accuracy against interpretability [19]. Critically, EFS are capable of inducing accurate and highly compact rule bases from the relatively small and imbalanced training datasets characteristic of rare diseases like NPC1, directly addressing the limitations previously highlighted.

While EFS offer a robust solution for static datasets, their deployment in healthcare is further complicated by the dynamic nature of clinical environments. Standard FL relies on a batch learning setting, assuming clients hold static, pre-collected datasets. This rarely holds in real-world deployments where new patient data arrives continuously and measurement distributions shift over time. Online or streaming machine learning mitigates this by enabling models to update incrementally [31]. The online variant of fuzzy systems, evolving fuzzy systems, extends this paradigm, allowing the rule base to be updated dynamically without full model retraining. However, the combination of FL with streaming data, termed online FL, remains critically underexplored. Initial explorations have studied FL under distributed concept

drift [32], proposed prototype-based approaches like FedStream [33], and investigated the interplay between drift detection and federated continual learning [34], finding that naive forgetting prevention mechanisms often fall short when distributions drift asynchronously.

Recognizing these converging needs, recent research has begun to explore the intersection of FL, EFS, and streaming data. For instance, Padilla-Rascon et al. [35] introduced TEFeS-SDR, a federated EFS designed to operate in a tabular data streaming environment, notably incorporating a security framework for the secure transmission of fuzzy rules.

The proposal presented in this work builds upon and transcends these foundational steps, constituting to the best of our knowledge, one of the first attempts to integrate online evolutionary fuzzy systems within a federated learning framework for rare disease applications. By integrating evolutionary optimization within a federated data online, the proposed algorithm not only preserves patient privacy and guarantees transparency through human-readable rules, but also facilitates continuous adaptation as multi-centric diagnostic registries update asynchronously. This targeted synergy overcomes the limitations of black-box models and static learning systems, providing a robust, scalable tool that satisfies both the strict regulatory demands for trustworthy AI and the urgent clinical need for early, accurate diagnosis in complex pathologies such as NPC1.

3 Online Federated Evolutionary Fuzzy System (OFEFS)

This section presents the Online Federated Evolutionary Fuzzy System (OFEFS) framework in detail, which constitutes the main methodological contribution of this study. OFEFS is proposed as a comprehensive solution that integrates a novel algorithm based on EFS with an online FL architecture. Its primary objective is the continuous and decentralized extraction of supervised descriptive rules to provide interpretable knowledge regarding the NPC1 disease. To address this dual design, the remainder of this section is divided into two main parts: Section 3.1 describes in depth the operators and internal mechanics of the newly implemented EFS algorithm, while Section 3.2 details the topology of the distributed architecture and its data flow.

3.1 Algorithm Description

The algorithm presented in this work, hereafter referred to as OFEFS, is an online rule-learning method that seamlessly combines three complementary paradigms: evolutionary computation, fuzzy logic, and FL. The core functionality of this algorithm relies on a specific representation of knowledge as a population of fuzzy rules, coupled with mechanisms by which that knowledge is incrementally refined and globally shared without compromising patient privacy.

At the foundation of this algorithm is the structural representation of the knowledge base. The model maintained by each client c_k is a fuzzy rule base $\mathcal{M}^{(k)} = \{r_1, r_2, \ldots, r_m\}$. Each rule r_i is formulated as a classification rule [36] whose antecedent

is a partial conjunction over a subset of the d input variables. Crucially, each condition within the antecedent may accept a disjunction of linguistic labels for the same variable. This allows a single rule to compactly cover multiple regions of the feature space along any given dimension. A rule therefore takes the mathematical form:

$$r_i : \text{IF } x_{j_1} \in \Lambda_{j_1}^i \text{ AND } \cdots \text{ AND } x_{j_s} \in \Lambda_{j_s}^i \text{ THEN } y = C_i$$

where $\{j_1, \ldots, j_s\} \subseteq \{1, \ldots, d\}$ is the variable-length set of active conditions of rule r_i, $\Lambda_j^i \subseteq \mathcal{L}_j^{(k)}$ is the non-empty subset of linguistic labels accepted by condition j in rule r_i, and $C_i \in \mathcal{Y} = \{\text{NPC1}, \text{Heterozygote}\}$ is the consequent class. For example, a rule expressing IF x_1 is A_2 AND x_3 is A_0 OR A_1 THEN class 0 dictates that the patient belongs to class 0 whenever variable x_1 takes label A_2 and variable x_3 takes either label A_0 or A_1. This representation is substantially more expressive than the traditional single-label-per-variable convention, as it empowers the evolutionary process to generalise conditions through label merging without artificially inflating the number of rules.

To operationalise this within an evolutionary framework, each rule r_i is treated as an individual within a population $\mathcal{P} = \mathcal{M}^{(k)}$, encoded by the genotype:

$$\mathbf{g}_i = \left\{ (x_j, \Lambda_j^i) \,:\, j \in S_i \right\} \cup \{C_i\}, \qquad \Lambda_j^i \subseteq \mathcal{L}_j^{(k)}, \quad S_i \subseteq \{1, \ldots, d\}$$

where S_i is the active condition set, which is structurally dynamic and may grow or shrink under genetic mutation. The linguistic labels in $\mathcal{L}_j^{(k)}$ underpinning these genes are defined by triangular membership functions [37], selected for their computational efficiency, interpretability, and well-understood behaviour in EFS [38]. A linguistic label A over a variable x_j is parameterised by a triple (a, b, c) with $a \leq b \leq c$, such that its membership degree is computed as:

$$\mu_A(x) = \max\left(0,\ \min\left(\frac{x-a}{b-a},\ \frac{c-x}{c-b}\right)\right)$$

where b is the modal value ($\mu_A(b) = 1$) and a, c represent the left and right feet of the triangle, respectively. Consequently, the overall activation degree of a rule r_i on a given observation $\mathbf{x}$ is computed as the product t-norm over its active conditions, where each condition contributes the maximum membership across its accepted labels:

$$\alpha_i(\mathbf{x}) = \prod_{j \in S_i} \max_{A \in \Lambda_j^i} \mu_A(x_j)$$

A rule r_i is considered to cover an observation $\mathbf{x}$ if $\alpha_i(\mathbf{x}) > 0$, and the entire model $\mathcal{M}^{(k)}$ successfully covers $\mathbf{x}$ if at least one of its constituent rules achieves a positive activation.

Beyond mere coverage, the evolutionary algorithm requires a robust mechanism to evaluate the quality of these rules. The fitness of a rule r_i is measured through a linear combination evaluated over a sliding window $\mathcal{W}^{(k)}$. The composite fitness function is

subsequently formulated as a configurable linear combination:

$$Q(r_i) = \sum_m \lambda_m \cdot f_m(r_i), \qquad \sum_m \lambda_m = 1, \quad \lambda_m \geq 0 \tag{1}$$

where each f_m corresponds to one of the aforementioned metrics, and the weight vector (λ_m) acts as a hyperparameter to control the weight of each metric. In this study, the weight vector (λ_m) is fixed uniformly ($\lambda_m = \frac{1}{3}$ for all components), reflecting an equal prioritization of confidence, discriminative capacity, and generality. While this choice simplifies the configuration, it is acknowledged that alternative weightings could influence rule selection and may be explored in future work through sensitivity analysis. This architectural design makes OFEFS highly adaptable to varied classification scenarios by simply adjusting the fitness landscape.

A central design principle dictating how this fitness evaluation is applied is the online nature of OFEFS. The algorithm deliberately avoids batch processing over static datasets; instead, it processes observations sequentially as they arrive, maintaining and refining a live model at each client node. This paradigm is essential in the clinical context of ultra-rare diseases like NPC1, where patient data accrues gradually over time across geographically distributed institutions, and waiting to collect a sufficiently large centralised dataset before training is neither practical nor privacy-preserving.

To manage this streaming environment efficiently, OFEFS adheres to a test-then-train protocol [31], which decouples model evaluation from the model update process, ensuring that computational effort scales with the novelty of the incoming stream rather than its total volume. Upon the receipt of a new patient observation $\mathbf{x}_t$, the system first evaluates it against the current model in a test phase. The evolutionary training cycle is triggered only if the model fails to satisfy two simultaneous quality criteria over the covered examples: a minimum confidence threshold θ_{conf} and a minimum true positive rate threshold θ_{tpr}. Formally, letting $\mathcal{A}(\mathbf{x}_t)$ be the set of rules currently covering $\mathbf{x}_t$, the training phase is initiated if:

$$\max_{r_i \in \mathcal{A}(\mathbf{x}_t)} \text{conf}(r_i) < \theta_{\text{conf}} \quad \vee \quad \max_{r_i \in \mathcal{A}(\mathbf{x}_t)} \text{tpr}(r_i) < \theta_{\text{tpr}}$$

This dual-threshold condition provides a highly discriminating trigger mechanism. A rule might formally cover an observation yet do so with low confidence or poor recall on its target class, both of which are highly undesirable outcomes in a medical diagnostic setting. If neither threshold is violated, the observation is simply incorporated into the local sliding window $\mathcal{W}^{(k)}$ and execution proceeds without triggering any model update or server communication, drastically minimising computational overhead during stable periods of the data stream.

When the algorithm determines that training is indeed necessary, it cannot rely on a single isolated observation. This limitation is resolved through the utilisation of the sliding window $\mathcal{W}^{(k)}$ of capacity w, which continuously retains the w most recent observations:

$$\mathcal{W}_t^{(k)} = \{\mathbf{x}_{t-w+1}, \mathbf{x}_{t-w+2}, \ldots, \mathbf{x}_t\}$$

This window provides a temporally coherent training corpus reflecting the recent distribution of the data stream, thereby enabling the evolutionary algorithm to induce a comprehensive rule base from a statistically meaningful sample. The capacity parameter w governs the well-known trade-off between temporal adaptability and statistical stability inherently present in online learning environments.

Provided with $\mathcal{W}_t^{(k)}$ and the current linguistic label set $\mathcal{L}^{(k)}$, OFEFS initiates a generational evolutionary cycle to produce an updated local model. The initial population is constructed randomly: each individual is assigned a random non-empty subset of variables as its active condition set, each condition receives a random non-empty subset of labels from the corresponding $\mathcal{L}_j^{(k)}$, and a random consequent class is drawn from $\mathcal{Y}$.

The evolutionary search then progresses through four distinct genetic operators. Selection is performed via tournament, where two individuals are drawn at random and the one exhibiting the higher fitness Q advances to the mating pool [39]. Crossover is subsequently applied to pairs of selected rules, strictly subjected to a consequent compatibility constraint. If the two parent rules carry conflicting class labels ($C_i \neq C_j$), the crossover operation is suppressed and both parents are returned as exact copies, guaranteeing the semantic coherence of the offspring. Conversely, when the consequents agree, a single cut-point is chosen uniformly at random over the union of the two parents' active condition sets; each offspring inherits a contiguous subset of conditions from each parent, with any duplicate conditions on the same variable being resolved by retaining one arbitrarily. Mutation is applied to each rule independently with a fixed probability, selecting uniformly among three structural alterations: (i) replacing the label subset of a randomly chosen active condition with a different subset from $\mathcal{L}_j^{(k)}$; (ii) adding a new condition on a currently inactive variable, provided the total number of active conditions does not exceed a predefined upper bound $s_{\max}$; or (iii) removing a randomly chosen active condition, provided at least one condition remains to maintain validity. These diverse operators ensure that the search process explores both discrete label assignments and overarching rule structures simultaneously. Finally, replacement is executed incorporating elitism; the n_e fittest individuals from the current generation are preserved unconditionally and injected into the subsequent generation, displacing the n_e worst offspring to guarantee the monotone improvement of the best solution across successive generations.

To further refine this evolutionary process in the presence of skewed data distributions, OFEFS incorporates two distinct execution modes. In the standard global mode, the evolutionary cycle runs over the entirety of the sliding window $\mathcal{W}^{(k)}$, inducing rules for all classes simultaneously. However, the system also supports a class-decomposition mode, wherein the cycle is executed independently once for each specific class $c \in \mathcal{Y}$. In this mode, the training set is strictly limited to the class-specific subset $\mathcal{W}_c^{(k)} = \{\mathbf{x} \in \mathcal{W}^{(k)} : y(\mathbf{x}) = c\}$ during each iteration. However, the fitness function $Q(r_i)$ remains structurally identical, and the model is evaluated on the entire window $\mathcal{W}^{(k)}$. The individual rule bases produced for each class are subsequently merged into a unified overarching model $\mathcal{M}^{(k)} = \bigcup_{c \in \mathcal{Y}} \mathcal{M}_c^{(k)}$, which is then subjected to the standard quality filtering and federated aggregation steps. This decomposition inherently follows the one-vs-rest principle and actively prevents the majority class

from dominating the evolutionary search space. This is a well-documented vulnerability when dealing with the severe class imbalance characteristic of NPC1 cohorts, and this approach ensures the generation of class-specific rule bases that remain robust, balanced, and highly interpretable from a clinical perspective.

3.2 System Architecture

The proposed system follows a federated client-server architecture in which a central server coordinates a set of n geographically distributed clients, $\mathcal{C} = \{c_1, \ldots, c_n\}$, each holding a private dataset $\mathcal{D}^{(k)}$ of patient records that is never shared outside the institution it belongs to. This design constitutes the foundational privacy-preserving mechanism of the system, as it ensures that sensitive clinical data remains local at all times, in accordance with the principles of FL [15].

The server fulfils two complementary aggregation responsibilities:

- On the one hand, it maintains a global model $\mathcal{M}_t^G$ by consolidating the local models $\{\mathcal{M}_t^{(k)}\}$ submitted by clients, accumulating the collective knowledge extracted across all institutions.
- On the other hand, it arbitrates the semantic consistency of the linguistic label sets $\{\mathcal{L}_t^{(k)}\}$ produced independently by each client, guaranteeing that all participants share a common fuzzy vocabulary before any rule comparison or aggregation takes place.

This linguistic alignment step is critical in a federated fuzzy setting: if clients operate with heterogeneous label granularities or incompatible membership function boundaries, the resulting rules become semantically incommensurable and cannot be meaningfully merged. The aggregation method applied to resolve this is described in [40].

Each client c_k operates an online EFS that incrementally updates its local model $\mathcal{M}_t^{(k)}$ as new patient data arrives. Clients are entirely independent of one another during local processing; their only interaction with the rest of the federation occurs through the server at well-defined synchronisation points. This independence is not merely a technical convenience but a deliberate design choice that supports scalability and resilience, a client can process its local stream and contribute to the global model regardless of the availability of the rest of the federation.

The system is not launched from a cold, uninformed state. Prior to any federated execution, a global seed model $\mathcal{M}_0^G$ is constructed from a single reference dataset of the same nature as those held by the clients. All clients receive this seed model as their initial local model $\mathcal{M}_0^{(k)}$, providing a meaningful starting point that accelerates convergence and ensures that even the earliest federated updates are grounded in prior domain knowledge. This warm-start strategy is especially relevant in rare disease scenarios such as NPC1, where the scarcity of labelled samples makes any a-priori information particularly valuable.

The runtime behaviour of each client, as illustrated by the sequence diagram in Figure 1, interleaves two concurrent processes: continuous local learning and periodic global synchronisation. Each time a new patient observation x_t arrives at client c_k, the

system first assesses whether the current model $M_t^{(k)}$ already covers that observation, that is, whether at least one existing fuzzy rule fires with sufficient activation. If coverage is confirmed, the observation is simply incorporated into the sliding window $W_t^{(k)}$, and execution continues without any communication overhead.

If, however, no existing rule covers $\mathbf{x}_t$, the system enters a full update cycle. Using the new observation together with the contents of $\mathcal{W}_t^{(k)}$, the client first re-derives its local linguistic label set $\mathcal{L}_t^{(k)}$ and submits it to the server. The server applies the aggregation procedure to produce a consensus label set $\bar{\mathcal{L}}_t$ that replaces the local labels of every participating client.

Immediately afterwards, the client induces a new local model $\mathcal{M}_t^{(k)}$ from the updated window and transmits it to the server. The server incorporates this model into its running aggregation, producing a new global model $\mathcal{M}_t^G$, which is then returned to the client. Not all rules in $\mathcal{M}_t^G$ are unconditionally accepted: only those whose quality measure exceeds a predefined threshold τ are merged into the local model, thus filtering out noisy or poorly supported rules that could degrade local performance. Finally, the observation $\mathbf{x}_t$ is added to $\mathcal{W}_t^{(k)}$, and the cycle resumes.

This two-level update mechanism, local adaptation followed by selective global integration, is what allows the system to simultaneously benefit from the statistical diversity of the federation and preserve the specificity of each client's local knowledge. Algorithm 1 formalises this process.

4 Experimental Study

This section presents the experimental study conducted to evaluate the performance and capabilities of the proposed OFEFS framework. To ensure a comprehensive and reproducible evaluation, the experiments and their corresponding analyses are structured into four main components. First, Section 4.1 describes the execution framework, detailing the technical environment and the distributed setup used for the simulations. Next, Section 4.2 introduces the NPC1 clinical datasets utilized in this study, along with the specific data preprocessing techniques applied to prepare the data streams. Following this, Section 4.3 outlines the hyperparameter configuration and algorithmic settings established for the proposed method. Finally, Section 4.4 presents and discusses the obtained results, evaluating both the performance of the model and the clinical interpretability of the extracted descriptive rules.

4.1 Experimental framework

To validate the OFEFS framework under realistic operational conditions, a physical federated testbed was assembled to simulate a multi-health center environment. The infrastructure follows a client-server topology distributed across the following hardware:

- Federated Clients (Health Centres (HC)): Three Raspberry Pi 4 Model B devices were deployed to act as independent clinical institutions. Each node is equipped with a Quad-core ARM Cortex-A72 CPU (1.5 GHz) and 4 GB of RAM. Each Raspberry

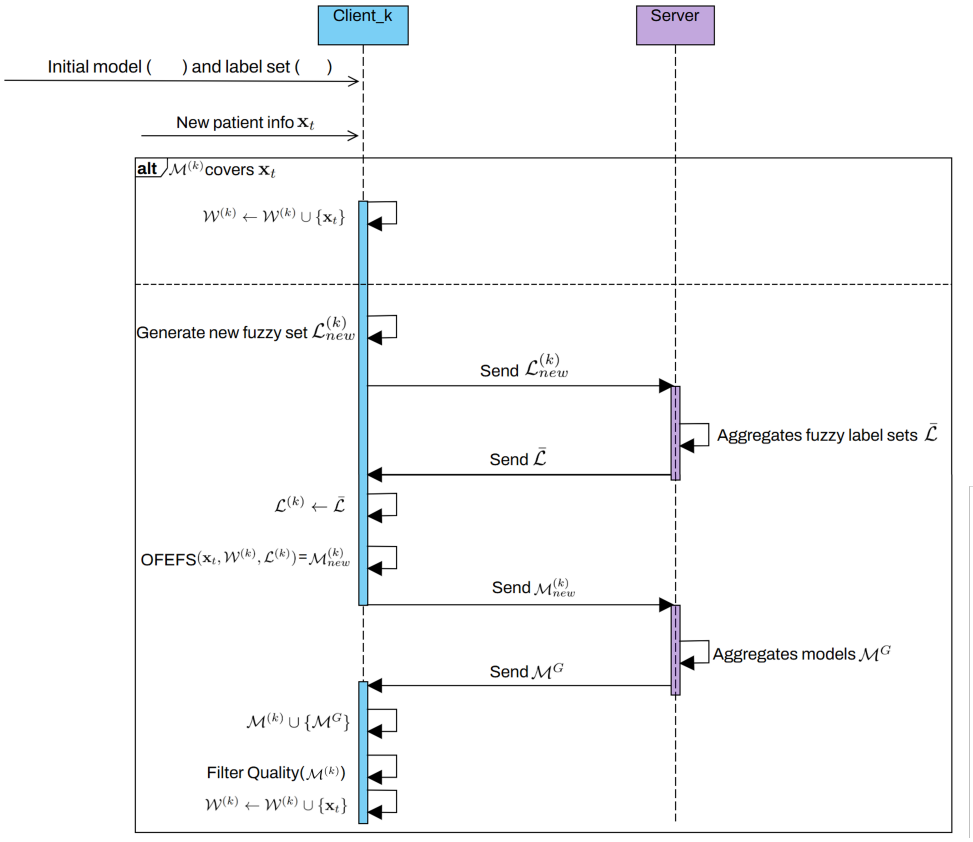


Fig. 1: Sequence diagram of the OFEFS incremental FL protocol. The workflow begins with a coverage check for new clinical instances; if an instance is not covered, a dual-stage federated synchronization is triggered: first, linguistic labels are locally updated and aggregated by the server, followed by the generation and global aggregation of local rule bases. The cycle concludes with the merging of local and global models, followed by a quality-based filtering stage to ensure a compact and interpretable rule set.

Pi operates autonomously, processing its own private patient dataset and extracting local fuzzy models.

- Central Server: A Linux-based server running Ubuntu 23.04 and powered by an Intel Core i7-1065G7 CPU (1.3 GHz base) with 8 GB of RAM coordinates the federation. The server is responsible for the dual task of linguistic label alignment and the aggregation of local models into a consolidated global knowledge base.

The conceptual architecture of the simulated environment is illustrated in Figure 2. This schema depicts the federation of three distinct HCs, each managing a heterogeneous local dataset. While the OFEFS framework and its hyper-parameters are identical across all nodes to ensure structural consistency, the resulting local models

Algorithm 1 Online Federated Evolutionary Fuzzy System (Client c_k)

Require: Seed global model $\mathcal{M}_0^G$, initial label set $\mathcal{L}_0^{(k)}$, sliding window $\mathcal{W}^{(k)}$ (max size w), quality threshold τ
Ensure: Continuously updated local model $\mathcal{M}^{(k)}$

$\mathcal{M}^{(k)} \leftarrow \mathcal{M}_0^G$ ▷ Warm-start from seed model
$\mathcal{L}^{(k)} \leftarrow \mathcal{L}_0^{(k)}$
while new observation $\mathbf{x}_t$ arrives **do**
  if $\mathcal{M}^{(k)}$ covers $\mathbf{x}_t$ **then** ▷ Rule fires above activation threshold
    $\mathcal{W}^{(k)} \leftarrow \mathcal{W}^{(k)} \cup \{\mathbf{x}_t\}$
    continue
  else
    ▷ Coverage failure: full update cycle begins
    $\mathcal{L}_{new}^{(k)} \leftarrow \text{InduceLabelSet}(\mathbf{x}_t, \mathcal{W}^{(k)})$
    Send $\mathcal{L}_{new}^{(k)}$ to server
    Receive consensus label set $\bar{\mathcal{L}}$ from server
    $\mathcal{L}^{(k)} \leftarrow \bar{\mathcal{L}}$ ▷ Replace local labels with consensus
    $\mathcal{M}_{new}^{(k)} \leftarrow \text{EFS-Train}(\mathbf{x}_t, \mathcal{W}^{(k)}, \mathcal{L}^{(k)})$
    Send $\mathcal{M}_{new}^{(k)}$ to server
    Receive global model $\mathcal{M}^G$ from server
    for each rule $r \in \mathcal{M}^G$ **do**
      if $Q(r) \geq \tau$ **then** ▷ Quality filter
        $\mathcal{M}^{(k)} \leftarrow \mathcal{M}^{(k)} \cup \{r\}$
      end if
    end for
    $\mathcal{W}^{(k)} \leftarrow \mathcal{W}^{(k)} \cup \{\mathbf{x}_t\}$
  end if
end while

diverge as they are incrementally trained on the specific data distributions of each institution. Positioned at the core of the architecture, the Aggregator Server facilitates the synchronization of linguistic labels and the consolidation of these diverse local insights into a unified global model.

4.2 Dataset and preprocessing

The foundation of this experimental framework relies on a UK-based clinical cohort comprising urine samples from 12 untreated Niemann-Pick type C1 (NPC1) disease patients (age range 2.7–30 years; 6 male, 6 female) and 40 corresponding heterozygous parental carrier controls (age range 25–62 years; 19 male, 28 female). All samples were collected with informed consent under Research Ethics Committee approval (06/MRE02/85) and immediately stored at -80°C to preserve metabolic integrity. High-resolution ^{1}H NMR spectra were subsequently acquired on a Bruker Avance AV-600 spectrometer operating at 600.13 MHz. To ensure spectral uniformity, preprocessing included Fourier transformation, phase and baseline correction, and intelligent

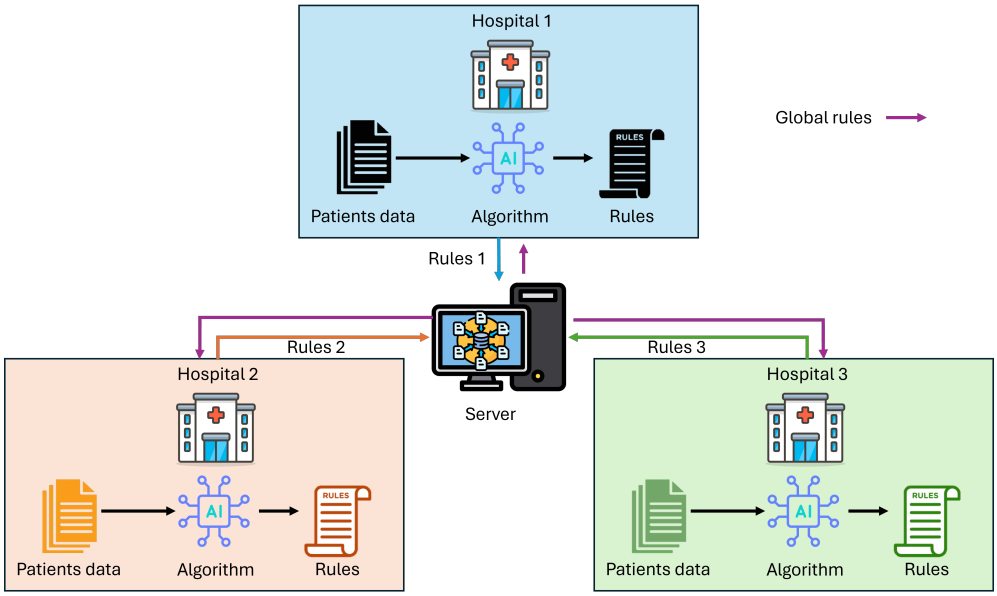


Fig. 2: Conceptual architecture of the OFEFS testbed. Each clinical institution (HC 1, 2, and 3) processes its private patient data locally to induce fuzzy rules, which are then synchronized and aggregated by a central server to form a global knowledge base.

bucketing (ISB) with a 0.04 ppm bucket width and a 50% looseness factor via ACD/Labs Spectrus Processor software. Following the removal of confounding resonances, specifically ethanol, urea, and residual H_2O/HOD, all metabolite intensities were normalized to the urinary creatinine N–CH_3 resonance ($\delta = 3.030$ ppm).

In its processed form, this dataset immediately presented a classical high-dimensionality challenge, commonly referred to as the $p \gg n$ problem . For rule induction algorithms, an exponentially growing search space drastically increases the risk of generating spurious or overspecific rules. To navigate this limitation effectively, Fuzzy Rough Feature Selection (FRFS) [41] was integrated into the pipeline prior to any model training. Grounded in fuzzy rough set theory, FRFS evaluates feature relevance without relying on strict linear correlation or distributional assumptions, making it exceptionally well-suited for clinical biomarker data characterized by continuous-valued attributes and overlapping class boundaries. By mathematically measuring how well candidate subsets preserve the positive region approximation quality of the full feature set with respect to the decision attribute, FRFS effectively condensed the original predictors into a highly informative, non-redundant vector of $d = 12$ features.

The impact of this selection is clear in Figure 3. The original 65-feature matrix (left) was heavily redundant, shown by the dense clusters of dark red and blue. After applying FRFS (right), the new 12-feature matrix is mostly uncorrelated, as shown by the shift toward lighter colors. This process effectively filtered out the noise, allowing the model to focus on a distinct set of predictors and reducing the risk of overfitting.

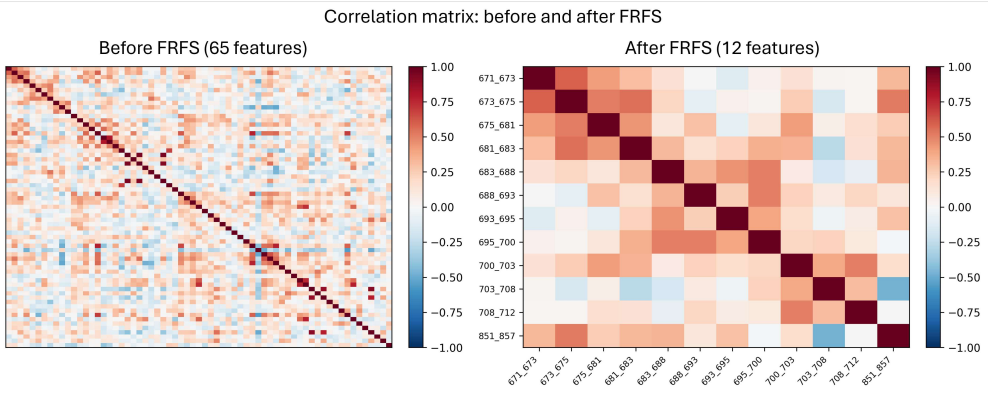


Fig. 3: Comparative correlation matrices demonstrating the dimensionality reduction efficacy of FRFS. (Left) The original high-dimensional feature space ($p = 65$), exhibiting severe multicollinearity and redundancy. (Right) The condensed feature space after FRFS application ($d = 12$), showing a significant reduction in inter-variable correlation, effectively mitigating the $p \gg n$ challenge prior to evolutionary rule induction.

Despite this optimized feature space, the dataset's intrinsic class imbalanc, a positive-to-negative ratio of approximately 1:3.3 reflecting the rarity of NPC1 [1], threatened to systematically bias the evolutionary search towards the majority class, degrading the model's sensitivity for the clinically critical minority class. To mitigate this, synthetic oversampling was applied using the NBalance algorithm [5]. Identified as the optimal data augmentation strategy for this specific cohort in preliminary comparative studies, NBalance generated representative synthetic minority-class instances until strict parity was achieved. This procedure yielded a stable, perfectly balanced working dataset comprising $N = 80$ instances, evenly split with 40 instances per target class.

The efficacy of this generative process is validated in Figure 4. As shown in the PCA projection, the synthetic NPC1 instances (green squares) are not randomly distributed; instead, they are strictly constrained within the manifold defined by the original clinical samples (blue crosses). This ensures that the augmented dataset preserves the local covariance structure of the metabolic profiles, effectively providing the evolutionary algorithm with a balanced training environment without introducing out-of-distribution artifacts that could lead to clinical misinterpretation.

Finally, to support the federated simulation framework, this balanced dataset was systematically partitioned into four disjoint subsets. The global seed dataset, denoted as $D^{(0)}$, was constructed via stratified random sampling to contain 50 instances (25 from each class). This foundational subset serves to initialize the shared global model $\mathcal{M}_0^G$, acting as a representative baseline akin to a publicly available reference cohort. The remaining 30 instances were distributed across three simulated HC clients $(D^{(1)}, D^{(2)}, D^{(3)})$ using the same stratified approach, preserving the approximate 1:1 class proportions within each local dataset. It is important to emphasize that the limited sample size could potentially introduce a constraint on the statistical power

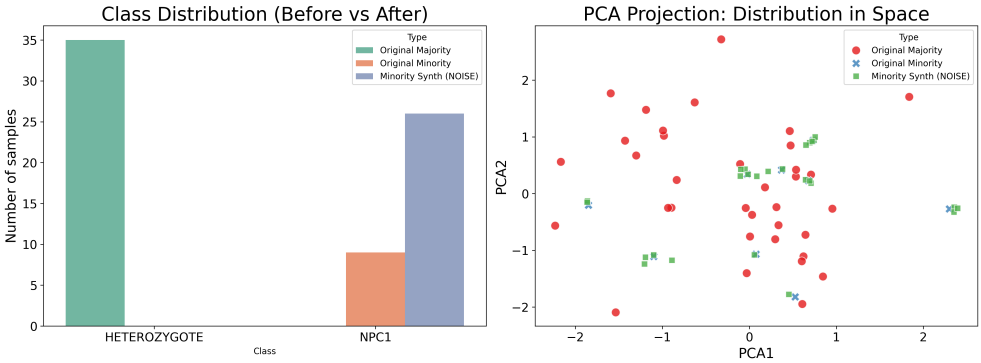


Fig. 4: Validation of the NBalance synthetic oversampling strategy. (Left) Class distributions before and after augmentation, showing the correction of the initial 1:3.3 imbalance to achieve strict dataset parity ($N = 80$). (Right) Principal Component Analysis (PCA) projection demonstrating that the synthetically generated NPC1 instances (green squares) accurately respect the localized topological distribution of the original minority patient samples (blue crosses) without encroaching upon the feature space of the majority carrier class (red circles).

and generalizability of the study. As a result, the findings should be interpreted as exploratory rather than definitive.

4.3 Algorithm Parameters

The configuration of the evolutionary algorithm was designed to balance the exploration of the search space with the latency constraints of the online environment. The population size was set to $|\mathcal{P}| = 100$ individuals with a maximum of $G = 50$ generations. These values were selected to avoid diminishing returns given the small scale of the local client datasets ($|D^{(k)}| \approx 10$). To mitigate the risk of premature convergence within such narrow data windows, exploration was prioritized through a crossover probability of $p_c = 0.80$ and a mutation probability of $p_m = 0.20$. Additionally, training triggers and the integration of global rules are governed by quality thresholds of $\theta_{\text{conf}} = 0.60$ and $\theta_{\text{tpr}} = 0.10$, while the sliding window capacity was fixed at $w = 4$. This relatively small window size reflects a deliberate trade-off between computational efficiency and adaptability in highly resource-constrained and data-scarce environments. However, it also limits the statistical robustness of local updates, and larger window sizes may yield more stable models when sufficient data is available.

Rule fitness is evaluated using a composite function $Q(r_i)$, which equally weights ($\lambda = 1/3$) confidence, support difference, and normalized Weighted Relative Accuracy ($WRAcc_{\text{norm}}$). While support difference penalizes rules that fail to discriminate adequately between classes, a critical factor in imbalanced clinical settings, the $WRAcc$ metric incentivizes rules that are simultaneously general and informative.

4.4 Results

The predictive performance of the global OFEFS model demonstrates a refined equilibrium between clinical reliability and diagnostic sensitivity. Evaluating the system through the lens of confidence, normalized Weighted Relative Accuracy ($WRAcc_{\text{norm}}$), and the True Positive Rate (TPR), the results suggest that the framework successfully synthesizes knowledge across distributed nodes without compromising the integrity of the induced rules. Specifically, the model achieves a high confidence level indicating that the generated fuzzy rules are remarkably consistent and provide a dependable basis for classification in a clinical setting. This reliability is further supported by a normalized $WRAcc_{norm}$ which confirms that the rules are not merely accurate but statistically meaningful, capturing non-trivial patterns within the metabolic profiles that would otherwise remain obscured. From a diagnostic perspective, the TPR reflects a robust capability to identify NPC1-positive patients. While there remains inherent difficulty in achieving perfect recall in such a rare and heterogeneous disease, this level of sensitivity is particularly significant given the federated and privacy-preserving constraints of the learning process. Taken together, these figures, which are summarized in Table 1, indicate that OFEFS successfully integrates knowledge across simulated system without sacrificing the interpretability or the clinical relevance of the resulting model.

Beyond these performance indicators, the clinical utility of OFEFS is deeply rooted in its structural transparency. Therefore, the interpretability of the global model is assessed by monitoring the total number of rules and the average number of variables per rule. A lower count in both metrics is preferred, as it minimizes the cognitive load on the clinician and ensures that each diagnostic suggestion can be traced back to a concise set of metabolic markers. The global model obtained through the federated aggregation of the three simulated HCs maintains a compact structure, avoiding the "curse of dimensionality" typically associated with $p \gg n$ problems. By keeping the antecedents brief and the rule base small, the system provides a clear window into the metabolic logic of NPC1, allowing experts to validate the findings against known biochemical pathways.

Table 1: Predictive performance and structural complexity metrics for the global OFEFS model.

# Rules	# Vars	Confidence	$WRAcc_{\text{norm}}$	TPR
35	1.89	0.8825	0.6539	0.7292

The qualitative nature of the learned rules provides perhaps the most direct evidence of the model's interpretability. A representative sample of the rules obtained is shown below in Rule set 1:

Rule 1: IF 700_703 = *low* $[-0.72495, -0.05724]$ **THEN NPC1**

Rule 2: IF 700_703 = *high* $[0.03691, 0.58784]$ **THEN HETEROZYGOTE**

Rule 3: IF 671_673 = *low* $[-0.36074, 0.27298]$ **THEN NPC1**

Rule 4: IF 703_708 = *medium* $[-0.96657, 0.81613]$ **THEN HETEROZYGOTE**

Rule 5: IF 693_695 = *medium* $[-0.4653, 0.3963]$ **AND** 700_703 = *low* $[-0.72495, -0.05724]$ **THEN NPC1**

Rule 6: IF 671_673 = *medium* $[0.27298, 0.90671]$ **AND** 695_700 = *medium* $[-0.23514, 0.73323]$ **THEN HETEROZYGOTE**

Rule Set 1: Rule set obtained by the OFEFS algorithm for NPC1 disease detection.

Each rule involves only two or three biomarker conditions, making its antecedent directly readable without any post-hoc explanation.

The clinical interpretability of the extracted rule set is substantially reinforced when the identified spectral variables are mapped onto their underlying metabolic pathways. As established via intelligent spectral bucketing (ISB) of the high-resolution ^{1}H NMR spectra, the variables selected by FRFS and subsequently encoded in the rule antecedents correspond to a coherent and biochemically meaningful set of urinary metabolites (Table 2). Of the twelve variables retained after dimensionality reduction, eight derive from kynurenine pathway metabolites, providing strong evidence for the marked chemopathological role of this route in NPC1 disease [12].

Table 2: ISB variable identities for the spectral regions appearing in the OFEFS rule antecedents.

Code	ISB (ppm)	Metabolite Identity
1	6.71–6.73	Xanthurenate-C8H
2	6.73–6.75	Xanthurenate-C8H
3	6.75–6.81	Kynurenine-Ar-C5 / 3-Hydroxykynurenine-ArC3H / 4-Aminobenzoate-ArC3/C5H / Picolinate-ArC2H
4	6.81–6.83	4-Aminobenzoate-C3/C5H
5	6.83–6.88	Kynurenine-Ar-C3 / 3-(3-Hydroxyphenyl)-3-hydroxypropanoate-C1/C6-CH
6	6.88–6.93	Tyrosine-C3 and C5H / 3-Hydroxykynurenine-ArC4H / 3-Hydroxyanthranilate-ArC6H
7	6.93–6.95	Phenol-ArC2/C6-CH
8	6.95–7.00	Phenol-ArC2/C6H / Xanthurenate-C5H
9	7.00–7.03	1-Methylhistidine-ImC5H
10	7.08–7.12	Histidine-ImC5H / 3-Methylhistidine-ImC5H
11	8.51–8.57	Quinolinate-ArC6H / Nicotinate-ArC3H / Picolinate-ArC5H

The biochemical context underlying the OFEFS rule antecedents is summarised in Figure 5, which maps the FRFS-selected spectral variables onto their corresponding metabolic pathways. As illustrated, the variables recurring most prominently across the induced rules belong to two functionally opposed branches of the kynurenine pathway: a neuroprotective arm, whose metabolites appear systematically as *low* in NPC1-class antecedents (Rules 1, 3, and 5), and a neurotoxic arm, whose disruption underlies the heterozygote-class rules (Rules 4 and 6).

The kynurenine pathway constitutes the principal hepatic route for tryptophan catabolism and serves as a primary site for nicotinamide adenine dinucleotide biosynthesis. Under conditions of chronic immune activation, a hallmark of lysosomal storage disorders, the enzyme indoleamine 2,3-dioxygenase (IDO1) is upregulated, diverting tryptophan flux away from serotonin and melatonin synthesis and towards the production of neuroactive and, in excess, neurotoxic downstream metabolites [6, 7]. This shift manifests as elevated quinolinate and 3-hydroxykynurenine, potent excitotoxins. alongside reduced neuroprotective kynurenate, a pattern previously established as a salient urinary signature of NPC1 [12]. Critically, elevated kynurenine-to-tryptophan ratios are especially pronounced in adults with neurodegenerative disorders relative to age-matched controls, reflecting a systemic neuroinflammatory burden [10, 11]. This pathway therefore provides a mechanistic link between the lysosomal cholesterol accumulation characteristic of NPC1, the resulting neuroinflammation, and the urinary metabolic deviations that the OFEFS rules capture.

Among the non-tryptophan features contributing to the rule antecedents, signals attributable to histidine metabolism are noteworthy. Histidine serves as the obligate precursor for histamine and carnosine biosynthesis and has been implicated in cognitive impairment and neuropsychiatric dysfunction; its urinary derivative 3-methylhistidine, generated by the post-translational methylation of actin and myosin, acts as a sensitive marker of skeletal muscle protein catabolism [13], consistent with the progressive neuromuscular deterioration observed in NPC1 patients. Additionally, spectral contributions from xanthurenate (XA), a downstream product of 3-hydroxy-L-kynurenine within a sub-branch of the kynurenine pathway, warrant particular attention. XA exerts neuroactive effects through a dedicated G-protein-coupled receptor expressed in brain neurons, modulating dopamine release and intercellular signalling in a manner distinct from, and in some respects antagonistic to, kynurenate [42]. Its capacity to form stable chelation complexes with transition metal ions, including iron and zinc, is of direct relevance to NPC1 pathophysiology: diminished XA availability promotes iron bioaccumulation and reactive oxygen species generation, contributing to the oxidative cascade and neurodegeneration associated with unesterified cholesterol accumulation in lysosomes [43, 44]. The balance between kynurenate and XA may thus serve as a homeostatic indicator of glutamatergic and dopaminergic tone within the CNS, and its dysregulation could represent one of the molecular bridges between peripheral lysosomal dysfunction and central neurological deterioration in NPC1.

Taken together, the linguistic rules produced by OFEFS do not merely reflect statistical co-occurrences; they encode, in a compact and human-readable form, the disruption of a coherent metabolic network. The fact that a clinician can read Rule

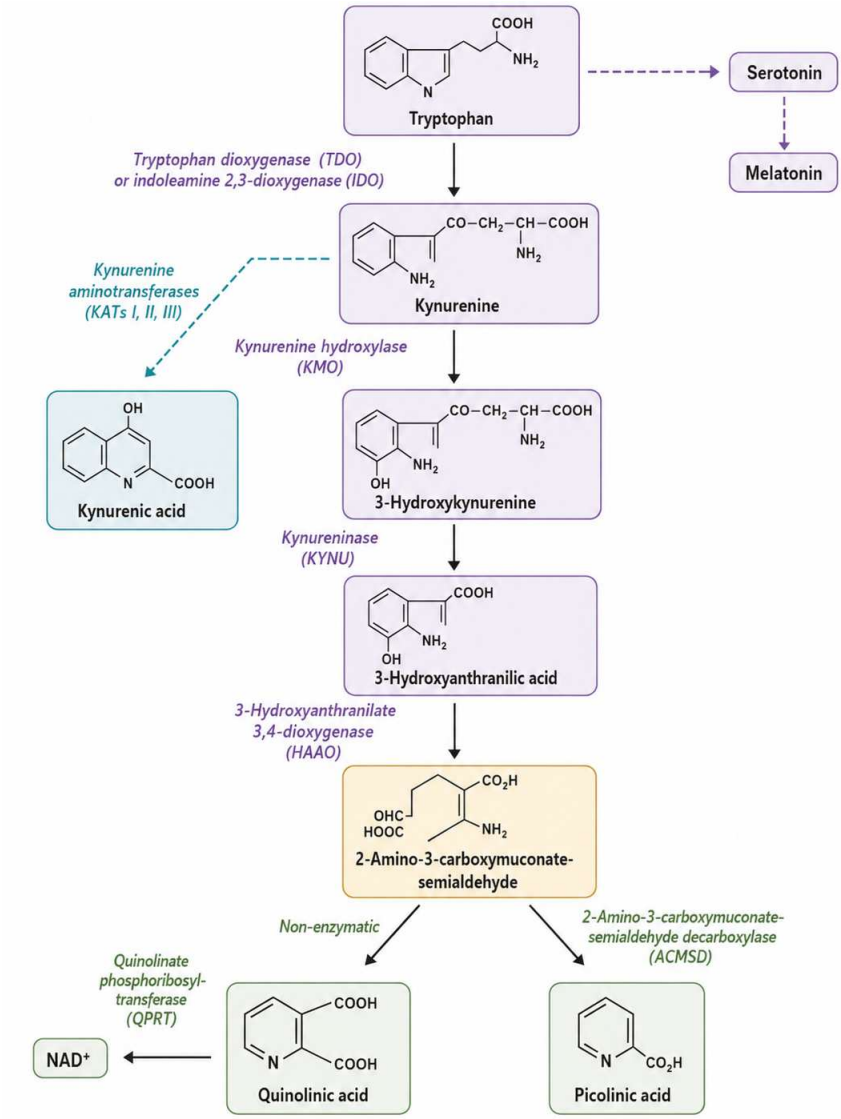


Fig. 5: Schematic of the kynurenine and histidine metabolic pathways implicated in NPC1, based on the variables selected by FRFS (Table 2). The diagram illustrates the metabolic shift from the neuroprotective branch (kynurenate, xanthurenate) toward the neurotoxic arm (quinolinate, 3-hydroxykynurenine) driven by IDO1/TDO upregulation. The histidine sub-pathway, relevant for heterozygote classification, is also depicted. Enzyme abbreviations: IDO1, TDO, KAT, KMO, HAAO, QPRT.

1: *IF 700_703 is low THEN NPC1* and trace it directly to reduced kynurenine-pathway metabolite excretion, or read Rule 5: *IF 693_695 is medium AND 700_703 is low THEN NPC1* and recognise the co-occurring perturbation in adjacent spectral regions linked to tryptophan catabolism, exemplifies the translational value of interpretable-by-design models in rare disease diagnostics. This biochemical coherence between the machine-induced rules and established NPC1 chemopathology constitutes independent, domain-level validation of the OFEFS framework, beyond what purely quantitative performance metrics can convey.

5 Conclusion

This paper has presented OFEFS, a federated online EFS designed for the detection of NPC1, a rare lysosomal storage disorder characterised by extreme data scarcity and the geographical dispersion of the few clinical cases that exist worldwide. The central contribution of the work is the integration of three paradigms that, whilst individually well-established, have not previously been combined in this form: FL, which ensures that sensitive patient data never leaves the institution that holds it; online incremental learning, which allows the system to adapt continuously as new patient observations arrive without requiring retraining from scratch; and evolutionary fuzzy rule induction, which produces models whose outputs are human-readable if-then rules directly auditable by clinicians, without recourse to post-hoc explanation methods.

The architectural design of OFEFS reflects a coherent response to the specific constraints of rare disease detection in a federated clinical setting. The test-then-train protocol limits computational overhead to periods of genuine model inadequacy, making the system viable on resource-constrained edge hardware. The dual-threshold triggering mechanism, the quality-filtered integration of globally aggregated rules, and the consensus-based alignment of linguistic label sets across clients collectively ensure that the federation produces a shared model that is both statistically grounded and semantically consistent. The warm-start initialisation from a seed model addresses the cold-start problem that would otherwise be acute given the small local datasets available at each client.

The experimental results suggest that OFEFS achieves a meaningful balance across the three quality dimensions that matter most in this context. The high rule confidence demonstrates that the learned rules are reliable when applied to new patients. The normalised WRAcc confirms that the rules are genuinely discriminative rather than trivially biased towards the majority class. The TPR, whilst leaving room for further improvement, reflects a clinically relevant sensitivity given the extreme scarcity of NPC1-positive training instances distributed across three independent clients. Critically, these quantitative results are accompanied by a compact and interpretable rule base whose individual rules can be reviewed and validated by a domain expert, a property that is not merely desirable but increasingly mandated by regulation for AI systems deployed in high-risk medical contexts, as codified in the EU AI Act.

The broader significance of this work lies in providing preliminary evidence that interpretability and federation can be jointly addressed within a unified framework.

OFEFS produces a globally shared model that is simultaneously the product of distributed private learning and a transparent, auditable clinical artifact. In a domain where both patient privacy and physician trust are non-negotiable, this combination is not a technical nicety but a prerequisite for deployment.

Furthermore, the biochemical coherence of the extracted rule base carries implications beyond diagnostic classification. The metabolic pathways encoded in the OFEFS rules, principally the kynurenine pathway and histidine metabolism, are themselves actionable therapeutic targets. The three major rate-limiting enzymes of the kynurenine pathway, namely indoleamine-2,3-dioxygenase 1 (IDO1), tryptophan-2,3-dioxygenase (TDO), and kynurenine monooxygenase (KMO), have received substantial attention as targets for intervention in neurodegeneration and cancer [7, 45]. The fact that OFEFS autonomously recovers rules grounded in precisely these pathways without any pathway-level supervision, suggests that the framework may generalise beyond NPC1 to other rare lysosomal storage disorders sharing similar metabolic disruptions, such as Fabry [8] and Gaucher's disease [9].

While the inherent scarcity of NPC1 data and the reliance on simulated environments represent initial constraints, they establish OFEFS as a novel proof-of-concept for federated clinical AI in rare disease domains. Acknowledging that the limited sample size and the necessary use of synthetic data augmentation characterize these findings as exploratory, future research will prioritize validation on larger, multi-center clinical cohorts alongside systematic benchmarking against alternative decentralized methods. Furthermore, the framework's architectural flexibility allows for its extension to multi-disease cohorts, investigating its capacity to facilitate cross-disease biomarker discovery and the differentiation between various lysosomal storage disorders sharing similar metabolic disruptions. Ultimately, this work serves as a scalable foundation for a global, privacy-preserving network dedicated to the advancement of pathway-targeted therapeutic strategies.

Acknowledgements. The research conducted in this study is supported by the FPU grant programme (FPU23/02620) and is part of the project "Advances in the Development of Trustworthy AI Models to Foster the Adoption and Use of Responsible AI in Healthcare (TAIH)" (PID2023-149511OB-I00), funded by the Spanish Ministry of Science, Innovation and Universities (MICIU) and the State Research Agency (AEI) (10.13039/501100011033), and co-funded by the European Regional Development Fund (ERDF), European Union.

Declarations

Funding

The research conducted in this study was supported by the FPU grant programme (FPU23/02620) and is part of the project "Advances in the Development of Trustworthy AI Models to Foster the Adoption and Use of Responsible AI in Healthcare (TAIH)" (PID2023-149511OB-I00), funded by the Spanish Ministry of Science, Innovation and Universities (MICIU) and the State Research Agency (AEI)

(10.13039/501100011033), and co-funded by the European Regional Development Fund (ERDF), European Union.

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

The data used in this study were collected with informed consent and previously approved by the appropriate Research Ethics Committee (06/MRE02/85).

Consent for publication

Not applicable.

Data availability

The datasets analysed during the current study are not publicly available due to privacy, institutional, and clinical restrictions, but may be made available from the corresponding author on reasonable request and subject to the appropriate approvals and authorisations.

Materials availability

Not applicable.

Code availability

The code used in this study is available from the corresponding author on reasonable request.

Author contributions

María Asunción Padilla-Rascón conceived the study, developed the methodology, designed and implemented the algorithmic architecture, conducted the experiments, analysed the results, and wrote the original manuscript. Martin Grootveld contributed the chemical expertise and domain-specific scientific content, and reviewed and edited the manuscript. Cristóbal J. Carmona, David Elizondo, and Ángel Miguel García-Vico contributed to manuscript review, supervision, and scientific discussion. All authors read and approved the final manuscript.

References

[1] Vanier, M.T.: Niemann-Pick disease type C. Orphanet J Rare Dis **5**(1) (2010) https://doi.org/10.1186/1750-1172-5-16 . Accessed 2026-04-13

[2] Geberhiwot, T., Moro, A., Dardis, A., Ramaswami, U., Sirrs, S., Marfa, M.P., Vanier, M.T., Walterfang, M., Bolton, S., Dawson, C., Héron, B., Stampfer, M., Imrie, J., Hendriksz, C., Gissen, P., Crushell, E., Coll, M.J., Nadjar, Y., Klünemann, H., Mengel, E., Hrebicek, M., Jones, S.A., Ory, D., Bembi, B., Patterson, M., International Niemann-Pick Disease Registry (INPDR): Consensus clinical management guidelines for Niemann-Pick disease type C. Orphanet J Rare Dis **13**(1) (2018) https://doi.org/10.1186/s13023-018-0785-7

[3] Cougnoux, A., Cluzeau, C., Mitra, S., Li, R., Williams, I., Burkert, K., Xu, X., Wassif, C.A., Zheng, W., Porter, F.D.: Necroptosis in Niemann-Pick disease, type C1: a potential therapeutic target. Cell Death Dis **7**(3) (2016) https://doi.org/10.1038/cddis.2016.16

[4] Patterson, M.C., Hendriksz, C.J., Walterfang, M., Sedel, F., Vanier, M.T., Wijburg, F., NP-C Guidelines Working Group: Recommendations for the diagnosis and management of Niemann-Pick disease type C: an update. Mol Genet Metab **106**(3) (2012) https://doi.org/10.1016/j.ymgme.2012.03.012

[5] Moreno-Barea, F.J., Franco, L., Elizondo, D., Grootveld, M.: Application of data augmentation techniques towards metabolomics. Computers in Biology and Medicine **148** (2022) https://doi.org/10.1016/j.compbiomed.2022.105916 . Accessed 2026-04-13

[6] Chen, Y., Guillemin, G.J.: Kynurenine pathway metabolites in humans: Disease and healthy states. International Journal of Tryptophan Research **2**, 2097 (2009)

[7] Platten, M., Nollen, E.A.A., Roehrig, U.F., Fallarino, F., Opitz, C.A.: Tryptophan metabolism as a common therapeutic target in cancer, neurodegeneration and beyond. Nature Reviews Drug Discovery **18**, 379–401 (2019)

[8] Uçar, T., Cansever, M.Ş., Isat, E., Zubairoğlu, T., Aktuğlu Zeybek, A.Ç., Topçu, B., Seyahi, N., Kıykım, E.: Serum neopterin, biopterin, tryptophan, and kynurenine levels in patients with Fabry disease. Balkan Medical Journal **41**(2), 113–120 (2024) https://doi.org/10.4274/balkanmedj.galenos.2024.2023-10-98

[9] Brautbar, A., Elstein, D., Pines, B., Krienen, N., Hemmer, J., Buskila, D., Zimran, A.: Fibromyalgia and Gaucher's disease. QJM: An International Journal of Medicine **99**(2), 103–107 (2005)

[10] Widner, B., Leblhuber, F., Fuchs, D.: Increased neopterin production and tryptophan degradation in advanced Parkinson's disease. Journal of Neural Transmission **109**, 181–189 (2002)

[11] Schwarcz, R., Rassoulpour, A., Wu, H.-Q., Medoff, D., Tamminga, C.A., Roberts, R.C.: Increased cortical kynurenate content in schizophrenia. Biological Psychiatry **50**, 521–530 (2001)

[12] Ruiz-Rodado, V., Marcos Luque-Baena, R., Te Vruchte, D., Probert, F., H. Lachmann, R., J. Hendriksz, C., E. Wraith, J., Imrie, J., Elizondo, D., Sillence, D., Clayton, P., M. Platt, F., Grootveld, M.: H NMR-Linked Urinary Metabolic Profiling of Niemann-Pick Class C1 (NPC1) Disease: Identification of Potential New Biomarkers using Correlated Component Regression (CCR) and Genetic Algorithm (GA) Analysis Strategies. CMB **02**(999) (2014) https://doi.org/10.2174/2213235X02666141013223224 . Accessed 2026-04-13

[13] Wishart, D.S., Feunang, Y.D., Marcu, A., Guo, A.C., Liang, K., Vazquez-Fresno, R., *et al.*: HMDB 4.0: The human metabolome database for 2018. Nucleic Acids Research **46**, 608–617 (2018)

[14] European Parliament and Council: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). Legislative Body: EP, CONSIL (2016). http://data.europa.eu/eli/reg/2016/679/oj Accessed 2026-04-13

[15] McMahan, B., Moore, E., Ramage, D., Hampson, S., Arcas, B.A.y.: Communication-Efficient Learning of Deep Networks from Decentralized Data. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. PMLR, ??? (2017). https://proceedings.mlr.press/v54/mcmahan17a.html Accessed 2026-04-13

[16] Hoi, S.C.H., Sahoo, D., Lu, J., Zhao, P.: Online learning: A comprehensive survey. Neurocomputing **459** (2021) https://doi.org/10.1016/j.neucom.2021.04.112 . Accessed 2026-04-13

[17] Gogineni, V.C., Werner, S., Gauthier, F., Huang, Y.-F., Kuh, A.: Personalized Online Federated Learning for IoT/CPS: Challenges and Future Directions. IEEE Internet Things M. **5**(4) (2022) https://doi.org/10.1109/IOTM.001.2200178 . Accessed 2026-04-13

[18] European Parliament and Council: Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance). Legislative Body: CONSIL, EP (2024). http://data.europa.eu/eli/reg/2024/1689/oj Accessed 2026-04-13

[19] Herrera, F.: Genetic fuzzy systems: taxonomy, current research trends and prospects. Evol. Intel. **1**(1) (2008) https://doi.org/10.1007/s12065-007-0001-5 . Accessed 2026-04-13

[20] Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F.: Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion **58** (2020) https://doi.org/10.1016/j.inffus.2019.12.012 . Accessed 2026-04-13

[21] Yang, G., Edwards, B., Bakas, S., Dou, Q., Xu, D., Li, X., Wang, W.: Federated learning as a catalyst for digital healthcare innovations. Patterns (N Y) **5**(7) (2024) https://doi.org/10.1016/j.patter.2024.101026

[22] Rauniyar, A., Hagos, D.H., Jha, D., Håkegård, J.E., Bagci, U., Rawat, D.B., Vlassov, V.: Federated Learning for Medical Applications: A Taxonomy, Current Trends, Challenges, and Future Research Directions. IEEE Internet of Things Journal **11**(5) (2024) https://doi.org/10.1109/JIOT.2023.3329061 . Accessed 2026-04-24

[23] Zhang, F., Zhai, D., Bai, G., Jiang, J., Ye, Q., Ji, X., Liu, X.: Towards fairness-aware and privacy-preserving enhanced collaborative learning for healthcare. Nat Commun **16**(1) (2025) https://doi.org/10.1038/s41467-025-58055-3 . Accessed 2026-04-24

[24] Casella, B., Riviera, W., Aldinucci, M., Menegaz, G.: MERGE: A model for multi-input biomedical federated learning. Patterns **4**(11) (2023) https://doi.org/10.1016/j.patter.2023.100856 . Accessed 2026-04-24

[25] Kairouz, P., McMahan, H.B.: Advances and Open Problems in Federated Learning. Foundations and Trends in Machine Learning **14**(1-2) (2021) https://doi.org/10.1561/2200000083 . Accessed 2026-04-24

[26] Moreno-Barea, F.J., Franco, L., Elizondo, D., Grootveld, M.: Data Augmentation Techniques to Improve Metabolomic Analysis in Niemann-Pick Type C Disease. In: Groen, D., Mulatier, C., Paszynski, M., Krzhizhanovskaya, V.V., Dongarra, J.J., Sloot, P.M.A. (eds.) Computational Science – ICCS 2022. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-08757-8_8

[27] European Commission: Ethics guidelines for trustworthy AI | Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai Accessed 2026-04-24

[28] Chaddad, A., Lu, Q., Li, J., Katib, Y., Kateb, R., Tanougast, C., Bouridane, A., Abdulkadir, A.: Explainable, Domain-Adaptive, and Federated Artificial Intelligence in Medicine. IEEE/CAA JAS **10**(4) (2023) https://doi.org/10.1109/JAS.2023.123123 . Accessed 2026-04-24

[29] Ducange, P., Marcelloni, F., Renda, A., Ruffini, F.: Federated Learning of XAI Models in Healthcare: A Case Study on Parkinson's Disease. Cogn Comput **16**(6) (2024) https://doi.org/10.1007/s12559-024-10332-x . Accessed 2026-04-24

[30] Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell **1**(5) (2019) https://doi.org/10.1038/s42256-019-0048-x . Accessed 2026-04-24

[31] Gama, J., Aguilar-Ruiz, J., Klinkenberg, R.: Knowledge discovery from data streams. Intelligent Data Analysis **12**(3) (2008) https://doi.org/10.3233/IDA-2008-12301 . Accessed 2026-04-24

[32] Jothimurugesan, E., Hsieh, K., Wang, J., Joshi, G., Gibbons, P.B.: Federated Learning under Distributed Concept Drift

[33] Mawuli, C.B., Che, L., Kumar, J., Din, S.U., Qin, Z., Yang, Q., Shao, J.: FedStream: Prototype-Based Federated Learning on Distributed Concept-Drifting Data Streams. IEEE Transactions on Systems, Man, and Cybernetics: Systems **53**(11) (2023) https://doi.org/10.1109/TSMC.2023.3293462 . Accessed 2026-04-24

[34] Casado, F.E., Lema, D., Criado, M.F., Iglesias, R., Regueiro, C.V., Barro, S.: Concept drift detection and adaptation for federated and continual learning. Multimed Tools Appl **81**(3) (2022) https://doi.org/10.1007/s11042-021-11219-x . Accessed 2026-04-24

[35] Padilla-Rascón, M.A., García-Vico, A.M., Carmona, C.J.: Trustworthy and explainable federated system for extracting descriptive rules in a data streaming environment. Results in Engineering **25** (2025) https://doi.org/10.1016/j.rineng.2025.104137 . Accessed 2026-04-24

[36] Mamdani, E.H., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. International Journal of Man-Machine Studies **7**(1) (1975) https://doi.org/10.1016/S0020-7373(75)80002-2 . Accessed 2026-04-24

[37] Zadeh, L.A.: Fuzzy sets. Information and Control **8**(3) (1965) https://doi.org/10.1016/S0019-9958(65)90241-X . Accessed 2026-04-24

[38] Cordón, O., Herrera, F., Hoffmann, F., Magdalena, L.: Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases. Advances in Fuzzy Systems — Applications and Theory, vol. 19. WORLD SCIENTIFIC, ??? (2001). https://doi.org/10.1142/4177 . https://www.worldscientific.com/worldscibooks/10.1142/4177 Accessed 2026-04-24

[39] Goldberg, D.E., Holland, J.H.: Genetic Algorithms and Machine Learning. Machine Learning **3**(2) (1988) https://doi.org/10.1023/A:1022602019183 . Accessed 2026-04-24

[40] Padilla-Rascón, M.A., García-Vico, A.M., Carmona, C.J.: Synchronising Fuzzy Labels in Federated Evolutionary Systems: A Preliminary Study. In: Martínez,

L., Camacho, D., Yin, H., Dutta, B., Yera, R., Rodríguez Domínguez, R.M., Tallón-Ballesteros, A. (eds.) Intelligent Data Engineering and Automated Learning – IDEAL 2025. Springer, Cham (2026). https://doi.org/10.1007/978-3-032-10489-2_21

[41] Jensen, R., Shen, Q.: Fuzzy-Rough Sets Assisted Attribute Selection. IEEE Transactions on Fuzzy Systems **15**(1) (2007) https://doi.org/10.1109/TFUZZ.2006.889761 . Accessed 2026-04-24

[42] Maitre, M., Taleb, O., Jeltsch-David, H., Klein, C., Mensah-Nyagan, A.-G.: Xanthurenic acid: A role in brain intercellular signaling. Journal of Neurochemistry **168**, 2303–2315 (2024)

[43] Kubicova, L., Hadacek, F., Bachmann, G., Weckwerth, W., Chobot, V.: Coordination complex formation and redox properties of kynurenic and xanthurenic acid can affect brain tissue homeodynamics. Antioxidants **8**, 476 (2019)

[44] Lovell, M.A., Robertson, J.D., Teesdale, W.J., Campbell, J.L., Markesbery, W.R.: Copper, iron and zinc in Alzheimer's disease senile plaques. Journal of the Neurological Sciences **158**, 47–52 (1998)

[45] Davis, I., Liu, A.: What is the tryptophan kynurenine pathway and why is it important to neurotherapy? Expert Review of Neurotherapeutics **15**(7), 719–721 (2015)