Evaluating Coalition Stability in Federated Learning Under Voluntary Client Participation

Abbas Zal, Thomas Marchioro, and Leonardo Badia
Department of Information Engineering, University of Padova, 35131 Padua, Italy
Email: {abbas.zal@studenti. thomas.marchioro@ leonardo.badia@} unipd.it

***Abstract*—Federated learning has become the de facto standard solution for decentralized training of machine learning models across multiple clients. However, when the participation of the clients is voluntary, the outcome can become unpredictable, especially if the quality of the data varies among the clients. In this work, we propose a game theoretic framework that models federated learning with voluntary participation as a coalition game. This framework determines stable client coalitions as NEs and predicts possible final outcomes in terms of federated model performance. We apply our proposed framework to public datasets commonly used for benchmarking federated learning solutions, empirically evaluating NEs. Our analysis investigates when a grand coalition of all clients is an equilibrium, which means that all clients have proper incentives to participate). When more than 40% of the clients have high-quality data, the grand coalition is an equilibrium approximately 90% of the time. However, when the proportion of low-quality clients increases, the number of cases where the full coalition is stable varies across different models.**

***Index Terms*—Federated learning; crowdsourcing; game theory; Nash equilibrium.**

I. Introduction

Federated learning (FL), introduced in [1] has become the main reference approach for decentralized machine learning, enabling multiple clients to collaboratively train a global model without sharing raw data. FL is particularly relevant in domains such as healthcare and finance [2]–[4], where data cannot be centralized due to legal or ethical constraints. In general, FL promises advantages in terms of privacy and data confidentiality [5].

Although initially limited by technical and infrastructural challenges, FL under voluntary client participation is becoming increasingly feasible due to the availability of open source frameworks and cloud-based platforms that facilitate collaboration [6]. However, client participation in FL is not without cost. Clients must allocate computational resources and bandwidth to communicate with the central aggregator. This may outweigh the benefits of participating if the global model does not outperform a locally trained model. This trade-off creates an incentive structure, where clients strategically decide whether to contribute their data to the global training process [7].

In this work, we introduce a game-theoretic framework that models federated learning under voluntary participation as a coalition game [8]. Game theory provides a powerful mathematical framework for analyzing strategic interactions in multi-agent systems, making it particularly well suited for studying coalition formation in federated learning, similar to its applications in crowdsourcing and other decentralized settings [9]. Our approach identifies stable client coalitions as Nash equilibria (NEs) and predicts possible outcomes in terms of the global model's performance.

The intersection of FL and game theory has initially been explored in many papers dealing with specific application scenarios, for the definition of proper incentives for nodes to participate in the FL system [10], with possible extensions to age of information for real-time traffic [11], or drones supporting an Internet-of-vehicle scenario [12]. From an abstract standpoint, the main game theoretic reference is that of coalitional games, which is also explored by many contributions. The foundation for this work can be related to [13], which analyzes coalitions in federated learning. The stability and optimality of coalitions was explored in a comparative fashion in a subsequent work by the same authors [14].

In many related papers [15], [16], this theoretical background is exploited in a simplified way, assuming that the so-called *grand coalition* is expected to form, i.e., all agents are willing to contribute and participate in training the same model, and whenever smaller coalitions are formed, they still contribute to the same global model.

In [17], it is assumed that the coalitions train their own unrelated FL model, so that data are only exchanged within the same coalition. However, our investigation still converges to the main point of this analysis in that we investigate the conditions for a grand coalition to be the preferred choice by clients.

In contrast, [18] and [19] explore mechanisms for efficient and stable coalition formation under dynamically changing conditions. The former considers a network of drones and proposed a hedonic coalition formation, whereas the latter is based on evolutionary mechanisms for fog computing where micro-providers are subject to variations on a fast timescale.

Unlike the aforementioned papers, who take a theoretical stance to the coalitional game approach to be effective, or define algorithms for coalition formation, we investigate a more practical perspective of real datasets to assess whether or not the coalition formation process is worth considering. We aim to evaluate whether the grand coalition is forming in practice and what is the role of data quality across the clients.

Indeed, many references make strong assumptions about the symmetry of data and the roles played by nodes, but under the assumption of voluntary network participation, the quality of data may be highly heterogeneous. As we show in this work, the grand coalition is generally forming in realistic scenarios, as long as the quality of the data from the sources is relatively similar. However, when the network consists mostly of nodes with low-quality data, the minority of sources whose data are good are generally not willing to deteriorate their model in a grand coalition. This possibly suggests interesting developments in vetting the quality of data before having a node join a coalition, to improve both stability of the resulting coalition and overall accuracy of the trained model.

To validate our claim, we study publicly available datasets commonly used to benchmark federated learning models. Our empirical analysis reveals that when at least 40% of the clients have high-quality data, every agent has a general incentive to contribute, so that the grand coalition is achieved as a Nash equilibrium in more than 90% of the cases. However, as the proportion of low-quality data holders increases, the stability of full participation becomes uncertain, depending on the specific learning model employed.

The paper is organized as follows. In Section II, we introduce the game theoretic model. Section III discusses the datasets that we used to validate the model. Section IV shows numerical results and Section V concludes the paper with a discussion of the results and future directions.

Notation. We denote a dataset of n samples with $\mathcal{D}=\{x_i, y_i\}_{i=1}^n$, where each sample $x_i \in \mathbb{R}^m$ is a vector of m features and y_i is a discrete label (since we focus on classification datasets). The total number of clients is denoted with N and $M \leq N$ is the size of coalitions.

Source code. Our results can be reproduced by using the code available at the following Github repository: https://github.com/abbaszal/FederatedLearningNE.

II. Game Theoretic Framework

We model federated learning under voluntary participation as a *static coalition game* with a finite set of players $\mathcal{N}$, where $|\mathcal{N}|=N$. Each player can decide whether to join (1) or not join (0) a coalition. The coalition structure is represented using binary encoding, where a binary string indicates the participation status of each player. For instance, the binary string '0000000011' means that players 1 and 2 have opted to join the coalition, while the others have not.

Utility Function

The utility of each player is based on the accuracy of the client's model. Specifically, we define the utility function as follows:

- Let $A_g(S)$ denote the *global accuracy* of the coalition $S \subseteq \mathcal{N}$.
- Let A_i denote the *local accuracy* of player $i \in \mathcal{N}$ when acting independently (i.e., without joining the coalition).

The utility function for each player $i \in \mathcal{N}$ is defined as:

$$U_i(S)=\begin{cases} A_g(S), & \text{if } i{\in}S \text{ (i.e., player joins the coalition)} \\ A_i, & \text{if } i{\notin}S \text{ (i.e., player remains independent)} \end{cases}$$

Furthermore, we assumed that if $A_g(S) = A_i$ player i prefers not to join the coalition, as participation offers no additional benefit. Notably, for simplicity, we neglect communication costs in this model.

NEs in FL with voluntary participation

A *Nash equilibrium* in this game occurs when no client has an incentive to unilaterally change their decision. For an NE to hold, *both* of the following conditions must be satisfied:

- Clients outside the coalition should not have an incentive to join, meaning their local accuracy satisfies $A_i \geq A_g(S)$.
- Clients within the coalition should not have an incentive to leave, which means they are not better off individually ($A_g(S) > A_i$).

If only one of these conditions holds, then the coalition is not at equilibrium, as some clients will still have an incentive to change their decision.

III. Models and Datasets

Datasets

We used two datasets that are widely adopted in federated learning studies.

Spambase [20]: Spambase is a dataset for spam email classification. It contains 4601 email instances, labeled as either spam or non-spam. The dataset includes $m = 57$ numerical features extracted from email content, categorized into three types: 48 word frequency features, 6 character frequency features, and 3 capital run length features (the latter quantify the use of capital letters, including average, longest sequence, and total count).

HuGaDB [21]: The Human Gait Database (HuGaDB) is a dataset designed for human gait analysis and activity recognition. It comprises continuous recordings of 12 distinct activities, including walking, running, sitting, standing, climbing stairs, etc. Data were collected from 18 healthy adults (4 women, 14 men) with an average age of 23.67 years, using a body sensor network. This network includes six wearable inertial sensors (3-axis accelerometers and gyroscopes) and two electromyography (EMG) sensors. Participants performed sequences of activities in continuous trials, resulting in approximately 2,111,962 samples, totaling about 10 hours of data. Each data file contains 39 columns: 36 for inertial sensor data, 2 for EMG data, and 1 for activity IDs. Since the dataset comprises multiple participants, its distribution presents an inherent multi-modal nature, which makes it a more realistic benchmark for federated learning applications [22].

Models

For the models, we utilized the federated version of two traditional machine learning models.

Federated Logistic Regression (FedLR) [23]. Each client i trains a logistic regression classifier of the form $y = \beta_i^T x + b_i, \beta_i \in \mathbb{R}^m, b_i \in \mathbb{R}$ in its local data set. Once local models are trained by the clients, their parameters are aggregated by the central server via federated averaging:

$$\beta = \frac{1}{M}\sum_{i=1}^{M}\beta_i, \quad b = \frac{1}{M}\sum_{i=1}^{M} b_i \tag{1}$$

with M being the number of participating clients.

Federated Forest (FedFor) [24]. In this case, each client trains a decision tree classifier on its local dataset. Since decision trees are non-parametric models, it is not possible to apply federated averaging. Instead, the local models are collected by the central server and their predictions are combined via majority voting:

$$\hat{y} = \arg\max_{c} \sum_{i=1}^{M} \mathbb{I}(y_i = c), \tag{2}$$

where y_i represents the prediction of the i-th tree, and $\mathbb{I}$ is the indicator function.

IV. Numerical Results

The following setup was used in our numerical experiments. For Spambase, we took an 80/20 random split between training and test data. The training data are further divided between clients, while the test data are kept as a single dataset and used to evaluate model performance. HuGaDB comes with default training and test splits for each user, which we adopt. The test data are grouped into a single dataset. However, due to its size, training was generally very slow. Since our experiments involve multiple training and coalition evaluation runs, we reduced the dataset size by subsampling 350 training samples for each client—matching the size of Spambase—and resampling them in each run.

The preprocessing applied to the datasets is as follows. Aside from the dataset-specific differences described above, the same preprocessing steps were applied to both datasets. As all features are numerical, standard scaling was applied using parameters (mean and standard deviation of each feature) computed from the training data of all users. These parameters can be estimated by clients exchanging aggregated statistics on their local datasets, as described in [1]. Labels were numerically encoded.

The models were trained using the following hyperparameter settings. For FedLR, the maximum number of iterations for each local model was set to 100. For FedFor, the maximum tree depth was limited to 100. These constraints were not the result of hyperparameter tuning but were chosen to limit training runtime, which can become expensive when performing multiple training and testing runs.

Each experimental run followed the same training protocol. In each run, the local client models were trained only once. This means that different coalitions were formed using the same set of trained models within a single run. However, the models were retrained between different runs, as the training data were resampled each time.

NE Search

To find NEs, we employed an exhaustive search over all non-empty and non-singleton coalitions, from 0000000011 (only the first two clients participate) to 1111111111 (*grand coalition*). At the NE, no client has incentive to unilaterally change their strategy. Therefore, to determine whether a coalition is an equilibrium, all the possible deviations of the N clients should be evaluated, as shown in Algorithm 1. If at least one client has incentive to deviate from their current strategy – meaning that doing so increases their accuracy – then the algorithm returns false. Otherwise, it returns true.

Algorithm 1 IsNashEquilibrium

Require: Coalition S, global accuracies A_g local accuracies A_i

for each client $i = 1, \ldots, N$ **do**
  if $i \in S$ **then**
    if $A_i \geq A_g(S)$ **then**
      return false
  else
    if $A_g(S \cup \{i\}) > A_i$ **then**
      return false
return true

Simulating clients with low-quality data

In our experiments, we study how participants with low-quality data affect NEs. To achieve this, we explored all configurations that involve the participation of all clients (both good-quality and low-quality clients). We then computed global model accuracies across different combinations, as well as local client accuracies. Based on these computations, we assessed the viability of stable (NE) coalitions for varying numbers of low-quality clients. The results of this analysis will be presented in the following.

To analyze how clients with low-quality data affect the NEs, we simulate their presence in a coalition through data corruption techniques, which are applied to some of the clients' local datasets. We call the clients whose data have been perturbed as "low-quality clients" (LQC). The corruption techniques involve perturbing entries with noise or entirely removing some values. More specifically, we choose a fraction p of the samples (x_i) to be perturbed. Of that fraction, a subset qp is perturbed with Gaussian noise $\mathcal{N}(0, \sigma^2)$. The remaining $(1-q)p$ samples have features replaced by null values (NaN).

Labels (y_i) are also perturbed by randomly flipping them to other classes with fixed probability r. This label corruption process is given by:

$$\hat{y}_i = \begin{cases} y_i, & \text{with probability } 1 - r, \\ y_i' \neq y_i, & \text{with probability } r. \end{cases} \tag{3}$$

Results

We first examine whether coalitions form NEs, which is verified by Algorithm 1, with a particular focus on the frequency of specific coalitions appearing as equilibria. We considered all four possible combinations of datasets and models, which leads to every plot or table entry being repeated four times. In particular, we are clearly interested in determining the features of the grand coalition in this sense.

Table I reports the occurrence of the grand coalition (as opposed to other coalitions) among the equilibria, and also quantifies their global accuracy for a network without LQCs. As a general trend, one can remark that the grand coalition is an equilibrium in the vast majority of cases, and it also generally achieves high accuracy.

Although this property generally holds, there are specific differences that are worth noting. First of all, the accuracy of HuGaDB is generally lower than that of Spambase. This happens because HuGaDB is a multiclass dataset, which is naturally more challenging than Spambase that only aims at binary classification. Of all the four combinations, it is interesting to notice how the FedFor-Spambase pair frequently results in scenarios where at least another coalition is an equilibrium in addition to the grand coalition. However, none of them has more than 4 occurrences.

This leads us to extend the analysis to the presence of a varying number of LQCs, which is done in the following figures.

TABLE I
NEs OCCURRENCES ON 100 TRIALS

	NE	# times	Glob. accuracy (std)
FedFor HuGaDB	Grand coalition	99	0.7552 (0.0306)
	Other coalitions	15	0.4432 (0.0665)
FedLR HuGaDB	Grand coalition	100	0.6641 (0.0093)
	Other coalitions	30	0.4592 (0.0148)
FedFor Spambase	Grand coalition	99	0.9203 (0.0085)
	Other coalitions	90	0.7532 (0.0859)
FedLR Spambase	Grand coalition	95	0.9277 (0.0036)
	Other coalitions	7	0.9153 (0.0079)

NE for varying LQC

The corruption parameters for the LQC were set as $p = 0.8$, $q = 0.5$, and $r = 0.2$. The noise standard deviation σ was varied to assess the impact of different noise levels on the NE.

Fig. 1 reports 4 plots for all possible combinations of datasets and methods, and shows the occurrence of the most frequently encountered NE. The labels above each bar display the mean global accuracy of the NE over all occurrences. It can be seen that FedFor substantially limits the frequency of coalitions not involving participation by all nodes. Conversely, FedLR Spambase has relatively higher occurrences of these coalitions, even surpassing the grand coalition if there are many LQCs. The frequency of the grand coalition appears to decrease generally in the number of LQCs.

Fig. 2 instead shows specifically the frequency of the grand coalition as an equilibrium, versus the number of LQCs. Here, we also consider a different value of σ. It can be seen that this frequency generally decreases when the number of LQC increases, but always raises again when the number of LQC is

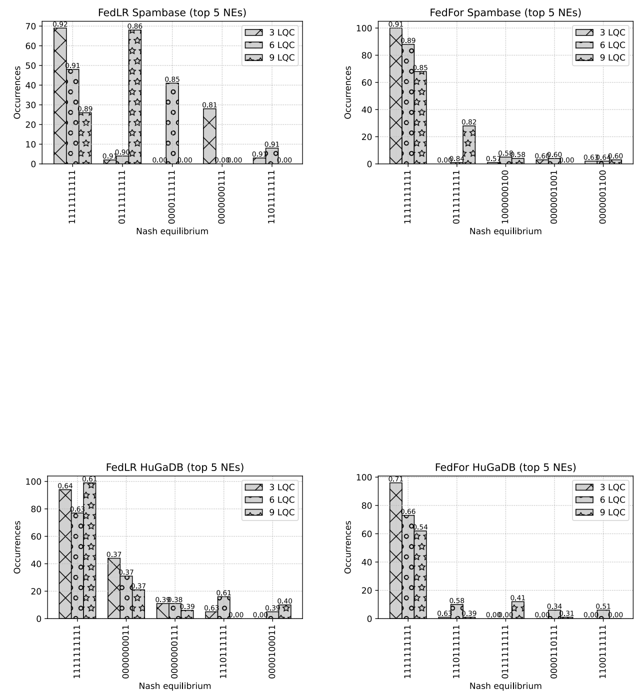

Fig. 1. Top 5 NEs (by frequency) for varying numbers of low-quality clients. Labels are the mean global accuracy across all occurrences.

so high that it approaches the totality of the network, which is quite natural, since the situation returns to being symmetrical.

More in general, the grand coalition tends to remain the most frequent NE under various conditions with varying numbers of low-quality clients. However, we observed a decrease in its frequency as the number of low-quality clients increased, which is expected.

Our empirical results show that the frequency of the grand coalition decreases with the number of low quality clients, and the trend becomes generally more pronounced if the standard deviation of the noise increases. This means that, for higher noise, i.e., when LQCs become worse, high-quality clients are less willing to join in a grand coalition. This trend is consistently confirmed across both datasets and models when the standard deviation of the noise is high enough. A notable exception is FedLR-Spambase, where the frequency is already lower than 100% when LQCs are a large fraction of the participants, even in the presence of low noise. Still, also in this case the occurrences of the grand coalition are reduced when the noise increases.

In general, the grand coalition is likely to form if the number of LQCs is limited, but starts to break when the fraction of LQCs becomes significant (more than half of the nodes). In

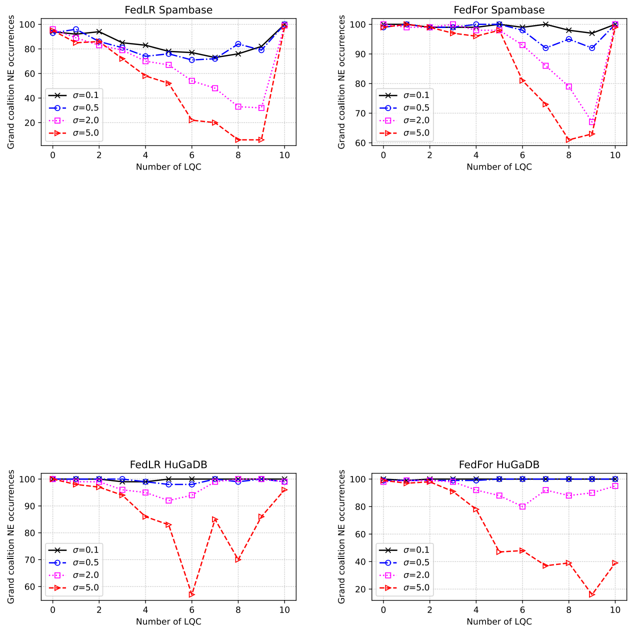

Fig. 2. Occurrences of the grand coalition equilibrium with varying LQC.

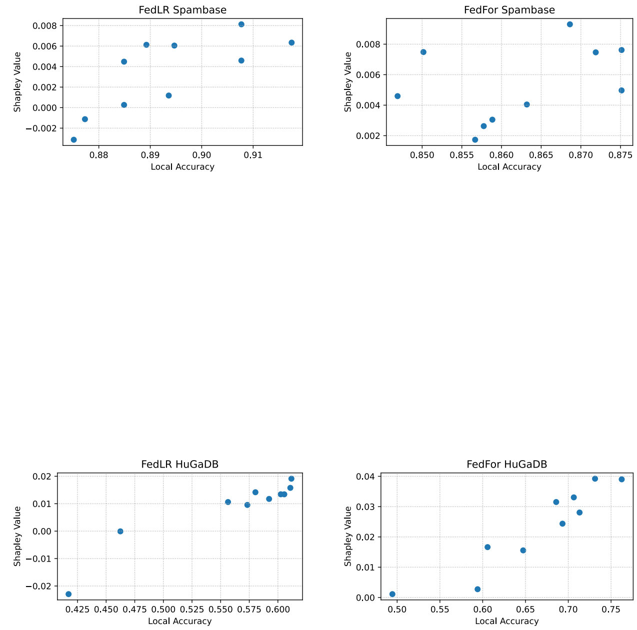

Fig. 3. Local accuracy versus Shapley value for each client.

this case, high-quality clients become less and less likely to join the grand coalition because it does not provide benefits compared to other coalitions. However, this decision is also influenced by the performance of the training model. When a model performed well, high-quality clients were more willing to participate.

In fact, clients are unable to make a perfect prediction about the outcome of each coalition. However, they may estimate the contribution of others based on their local accuracy. A typical way to represent the individual contribution of a client in a coalition is represented by the Shapley value [25], which is computed for client i in coalition $\mathcal{S}$ as

$$\phi_i = \sum_{\substack{\mathcal{S} \subseteq \mathcal{N}: \\ i \in \mathcal{S}}} \left[\frac{(|\mathcal{S}|-1)!\,(|\mathcal{N}|-|\mathcal{S}|)!}{|\mathcal{N}|!} \Big(A_g(\mathcal{S}) - A_g(\mathcal{S} \setminus \{i\}) \Big) \right].$$

The formula computes the average marginal contribution of i to the accuracy, considering differences in the predictions of the model with and without i combined over the possibilities of forming $\mathcal{S}$ [26].

The experimental results show, for all four scenarios, a generally good level of correlation between local accuracy and the Shapley value of the client, as seen in Fig. 3. Clients with poor global test performance consistently received low Shapley values, demonstrating that weak models or low-quality data contribute minimally. Although a client's highest local accuracy did not always yield the highest Shapley value, potentially due to over-specialization, we can conclude that local accuracy generally proved to be a reliable predictor of Shapley values.

V. Conclusions

Summary of Findings

The grand coalition frequently emerged as an equilibrium, indicating that full participation benefits collaborative learning and incentivizes even high-quality clients. In the HuGaDB dataset, alternative equilibria appeared. When excluding the grand coalition ('1111111111'), FedLR typically led to a mix of high- and low-quality clients, whereas FedFor mostly resulted in NEs with predominantly low-quality clients.

Our analysis shows that grand coalitions are frequently ($\sim$ 90% of the times) an equilibrium when more than 40% of the clients have high-quality data. This means that all clients have an incentive to participate in normal conditions [10]. When a portion of the clients have low-quality data, the frequency of the grand coalition as an NE is reduced, reaching its minimum when $N-1$ clients own low-quality data.

Limitations

Our current model is subject to a number of limitations and challenges. First, the assumption of complete information may not reflect the clients' knowledge, and uncertainty should be incorporated into the model. This can be achieved by expanding the analysis to lotteries or Bayesian games [27]. These frameworks do not assume that clients know the exact value of their utility function, which implies knowing A_g. Instead, clients are assumed to estimate an expected value of their utility based on a prior distribution. Additionally, experiments on more realistic neural network models were unfeasible due to the complexity of NE search, which requires to evaluate all the possible coalitions.

Practical insights

In cases where the grand coalition is an equilibrium, clients have an inherent incentive to participate in FL. Analyzing the likelihood of this equilibrium can affect the design of incentive mechanisms for FL [28]. For example, in cases where the grand coalition occurs with certainty, clients do not need further rewards. Conversely, when the grand coalition is less likely, appropriate rewards can be estimated via a PoA-based analysis [29].

Future Work

Future research could extend these experiments to larger and more diverse data sets and validate the findings using additional models. Further studies could also extend the analysis to neural networks by evaluating approximate NEs, and, for example, using Monte Carlo simulations.

REFERENCES

[1] B. McMahan and D. Ramage, "Federated learning: Collaborative machine learning without centralized training data," 2017.

[2] R. S. Antunes, C. André da Costa, A. Küderle, I. A. Yari, and B. Eskofier, "Federated learning for healthcare: Systematic review and architecture proposal," *ACM Trans. Intell. Syst. Techn. (TIST)*, vol. 13, no. 4, pp. 1–23, 2022.

[3] F. Biancalani, G. Gnecco, A. Signori, and S. Vismara, "Financing decisions following negative shocks in the product market: A matrix-completion study of the us pharmaceutical industry," *Econ. Lett.*, vol. 243, p. 111936, 2024.

[4] P. D. Prasad, Y. Vivek, and V. Ravi, "FedPNN: One-shot federated classifier to predict credit card fraud and bankruptcy in banks," in *Proc. IEEE Symp. Comput. Intell. Financial Engin. Economics (CIFEr)*, 2024, pp. 1–8.

[5] T. Marchioro, L. Giaretta, E. Markatos, and S. Girdzijauskas, "Federated naive Bayes under differential privacy," in *Proc. Int. Conf. Sec. Cryptog. (SECRYPT)*, 2022, pp. 170–180.

[6] A. Ziller, A. Trask, A. Lopardo, B. Szymkow, B. Wagner, E. Bluemke, J.-M. Nounahon, J. Passerat-Palmbach, K. Prakash, N. Rose *et al.*, "Pysyft: A library for easy federated learning," *Federated learning systems: Towards next-generation AI*, pp. 111–139, 2021.

[7] S. R. Pandey, N. H. Tran, M. Bennis, Y. K. Tun, Z. Han, and C. S. Hong, "Incentivize to build: A crowdsourcing framework for federated learning," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2019.

[8] P. P. Shenoy, "On coalition formation: a game-theoretical approach," *Int. J. Game Th.*, vol. 8, no. 3, pp. 133–164, 1979.

[9] L. Badia and N. Borra, "Crowdsensing strategies inspired by choir management analyzed via game theory," in *Proc. IEEE Ann. Ubiquitous Comp. El. Mob. Commun. Conf. (UEMCON)*, 2020, pp. 0862–0868.

[10] M. Cong, H. Yu, X. Weng, and S. M. Yiu, "A game-theoretic framework for incentive mechanism design in federated learning," *Feder. Learning: Privacy Incent.*, pp. 205–222, 2020.

[11] A. Buratto, A. Mora, A. Bujari, and L. Badia, "Game theoretic analysis of AoI efficiency for participatory and federated data ecosystems," in *Proc. IEEE Int. Conf. Commun. (ICC) Wkshps*, 2023, pp. 1301–1306.

[12] J. S. Ng, W. Y. B. Lim, H.-N. Dai, Z. Xiong, J. Huang, D. Niyato, X.-S. Hua, C. Leung, and C. Miao, "Joint auction-coalition formation framework for communication-efficient federated learning in UAV-enabled Internet of vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 4, pp. 2326–2344, 2020.

[13] K. Donahue and J. Kleinberg, "Model-sharing games: Analyzing federated learning under voluntary participation," https://arxiv.org/abs/2010.00753, 2020, arXiv 2010.00753.

[14] ——, "Optimality and stability in federated learning: A game-theoretic approach," *Adv. Neur. Inf. Proc. Syst.*, vol. 34, pp. 1287–1298, 2021.

[15] S. Jiang and J. Wu, "Coalition formation game in the cross-silo federated learning system," in *Proc. IEEE Int. Conf. Mob. Ad Hoc Smart Syst. (MASS)*, 2022, pp. 49–57.

[16] A. Buratto, E. Guerra, M. Miozzo, P. Dini, and L. Badia, "Energy minimization for participatory federated learning in IoT analyzed via game theory," in *Proc. Int. Conf. Artif. Intell. Inf. Commun. (ICAIIC)*, 2024, pp. 249–254.

[17] M. Ota, Y. Sakurai, and S. Oyama, "Coalitional game theoretic federated learning," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Techn. (WI-IAT)*, 2022, pp. 48–55.

[18] J. Liu, X. Li, Y. Xu, C. Lyu, Y. Wang, and X. Liu, "Hedonic coalition formation game and contract-based federated learning in UAV-assisted Internet of things," *IEEE Internet Things J.*, 2025, to appear.

[19] Z. Yasser, A. Hammoud, A. Mourad, H. Otrok, Z. Dziong, and M. Guizani, "Federated learning and evolutionary game model for fog federation formation," *IEEE Internet Things J.*, vol. 12, no. 4, pp. 4183–4196, 2025.

[20] M. Hopkins, E. Reeber, G. Forman, and J. Suermondt, "Spambase [dataset]," UCI Machine Learning Repository, https://doi.org/10.24432/C53G6X, 1999.

[21] R. Chereshnev and A. Kertész-Farkas, "HuGaDB: Human gait database for activity recognition from wearable inertial sensor networks," in *Proc. Anal. Images Social Netw. Texts. (AIST)*, 2017, pp. 120–130.

[22] M. Favero, T. Marchioro, and C. Schiavo, "Federated forests with differential privacy for distributed wearable sensors," in *Proc. IEEE Int. Conf. Aerospace Elec. Remote Sensing Techn. (ICARES)*, 2024, pp. 1–6.

[23] T. Yang, G. Andrew, H. Eichner, H. Sun, W. Li, N. Kong, D. Ramage, and F. Beaufays, "Applied federated learning: Improving google keyboard query suggestions," *arXiv preprint arXiv:1812.02903*, 2018.

[24] Y. Liu, Y. Liu, Z. Liu, Y. Liang, C. Meng, J. Zhang, and Y. Zheng, "Federated forest," *IEEE Trans. Big Data*, vol. 8, no. 3, pp. 843–854, 2020.

[25] L. S. Shapley, "A value for n-person games," *Contrib. Th. Games II, Ann. Math. Stud.*, vol. 28, 1953.

[26] E. Borella, U. B. Çakmakçi, E. Gottardis, A. Buratto, T. Marchioro, and L. Badia, "Effective sensor selection for human activity recognition via Shapley value," in *Proc. IEEE Int. Wkshp Metrology Living Envir. (MetroLivEnv)*, 2024, pp. 22–27.

[27] A. Tolio, D. Boem, T. Marchioro, and L. Badia, "Spreading factor allocation in LoRa networks through a game theoretic approach," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2020, pp. 1–6.

[28] L. Badia, M. Levorato, F. Librino, and M. Zorzi, "Cooperation techniques for wireless systems from a networking perspective," *IEEE Wirel. Commun.*, vol. 17, no. 2, pp. 89–96, 2010.

[29] V. Mancuso, P. Castagno, L. Badia, M. Sereno, and M. Ajmone Marsan, "Optimal allocation of tasks to networked computing facilities," in *Proc. Int. Conf. Anal. Stoch. Model. Techn. Appl. (ASMTA)*, 2024, pp. 33–50.