REPUBLIC OF AZERBAIJAN

On the rights of the manuscript

ABSTRACT

of the dissertation for the degree of Doctor of Philosophy

ALTERNATIVE METHODS FOR AGGREGATION IN FEDERATED LEARNING

Speciality: 3338.01- System analysis, control and information processing (control and decision making)
Field of science: Technical Sciences
Applicant: **Samir Ilgar Aliyev**

Baku-2026

The work was performed at "General and applied mathematics" departments of Azerbaijan State Oil and Industry University.

Scientific Supervisors:	Honorary scientist, Doctor of mathematical sciences, prof. **Araz Rafig Aliyev**
	Doctor of Philosophy on engineering, assoc. prof. **Nigar Tahir Ismayilova**
Official Opponents:	Doctor of technical sciences, prof. **Fazil Hazin Alakbarli**
	Doctor of technical sciences, assoc. prof. **Yadigar Nasib Imamverdiyev**
	Doctor of Philosophy on engineering, assoc. prof. **Sevinj Hamzaga Aliyeva**

Dissertation council FD 2.48 of Supreme Attestation Commission under the President of the Republic of Azerbaijan operating at Azerbaijan State Oil and Industry University

Chairman of the Dissertation council: ____________________	Corresponding member of ANAS, Doctor of technical sciences, prof. **Rafig Aziz Aliyev**
Scientific secretary of the Dissertation council: ____________________	Doctor of technical sciences, assoc. prof. **Akif Vali Alizadeh**
Chairman of the Scientific seminar: ____________________	Doctor of technical sciences, prof. **Kamala Rafig Aliyeva**

GENERAL DESCRIPTION OF THE DISSERTATION

Relevance of the research topic and degree of elaboration. Machine Learning is one of the fastest-growing fields of the modern era and plays a crucial role across a wide range of domains, from scientific research to business, healthcare, and engineering. Owing to the exponential growth of data volumes and advancements in computational power, machine learning enables the effective solution of problems that are difficult to address using traditional methods. By analyzing data, machine learning algorithms identify patterns and make decisions based on learned representations.

The relevance of machine learning is closely associated with its extensive application in automated decision-making processes. In particular, it is widely used to extract meaningful information from Big Data, generate predictions in complex systems, and minimize human intervention. Its impact is evident in fraud detection in the banking sector, early disease diagnosis in healthcare, optimization of industrial equipment performance, and the development of personalized recommendation systems.

As a fundamental branch of artificial intelligence, machine learning not only enhances existing technologies but also stimulates the emergence of new innovations. Through improved algorithmic accuracy and efficiency, data-driven decision-making, and advanced management of complex systems, machine learning has become an indispensable tool in the contemporary technological landscape. Its development contributes to increased productivity, optimization of human resource allocation, and the creation of intelligent systems.

Federated Learning, emerging alongside the advancement of machine learning, represents a significant approach whose relevance is primarily associated with privacy preservation, data security, and optimization of computational resources. Traditional machine learning models rely on transferring large volumes of data to centralized servers for processing. However, this centralized approach introduces practical challenges, particularly concerning privacy risks, as the transmission of sensitive data to central servers increases the likelihood of data breaches. Furthermore, transferring large datasets leads to network congestion and additional computational costs.

Federated Learning addresses these challenges by enabling decentralized model training directly on users' devices or local servers. In this paradigm, raw data remain on local devices, thereby preserving user privacy while allowing collaborative model training. This approach has broad applications in healthcare, banking, smart devices, and other domains. For example, hospitals can analyze patient data collaboratively without sharing raw records, and mobile devices can optimize personalized models locally.

Object and subject of the research. The object of the research is the coordinated training and aggregation mechanisms of models among clients within distributed federated learning environments aimed at preserving data privacy.

The subject of the research is the analysis and development of client selection and weighting methods to ensure reliable learning in federated learning systems. Specifically, the study investigates the impact of Fuzzy Inference Systems (FIS), the Analytical Hierarchy Process (AHP), Genetic Algorithms, and Coordinate Descent methods on model training quality, convergence, and generalization performance.

Aim and objectives of the research

The main objective of the dissertation is to improve aggregation methods in federated learning environments through the application of differentiated weighting strategies and adaptive weighting mechanisms based on fuzzy modeling. The study also investigates the limitations of existing aggregation techniques and proposes novel personalized weighting strategies evaluated across multiple datasets.

- To achieve this objective, the following tasks have been defined:
- To analyze aggregation methods in existing federated learning algorithms and identify their limitations;
- To develop novel weighting approaches that account for data heterogeneity and computational capabilities of different clients;

- To apply the Newton Direction method in federated learning and conduct a comparative analysis with the Federated Stochastic Gradient Descent (FedSGD) algorithm;
- To evaluate the impact of various weighting strategies within the Federated Personalization (FedPER) algorithm;
- To design and implement fuzzy logic-based weighting mechanisms;
- To develop an automatic rule selection mechanism for fuzzy logic systems using genetic algorithms;
- To experimentally evaluate the effectiveness of the proposed methods across different federated learning scenarios.

Research methods

To accomplish the research objectives, Fuzzy Inference Systems, decision-making theory, genetic algorithms, optimization methods, and machine learning techniques were employed. These approaches facilitated the improvement of weighting mechanisms in federated learning, enabled the development of adaptive methods that consider client heterogeneity, and enhanced the efficiency of aggregation processes.Müdafiəyə çıxarılan əsas müddəalar.

Main provisions submitted for defense

- The principal findings submitted for defense are as follows:
- The applicability of Fuzzy Inference Systems, the Analytical Hierarchy Process, and Genetic Algorithms in federated learning has been investigated, particularly their role in optimizing weighting mechanisms and enabling personalized approaches;
- The influence of differentiated weighting strategies on the Federated Averaging (FedAvg) algorithm has been analyzed;
- The effectiveness of various weighting strategies within the FedPER algorithm has been evaluated;
- A Federated Newton Direction (FedND) approach has been proposed and experimentally validated, demonstrating improved reliability compared to FedSGD;
- The impact of various weighting approaches on the

performance of distributed federated learning algorithms has been examined;

- Genetic algorithms have been applied for automatic rule selection in Fuzzy Inference Systems in expert-free environments.

Scientific novelty

The scientific novelty of the research is reflected in the following contributions:

- ✓ The application of FIS, AHP, Coordinate Descent, and Genetic Algorithms as weighting mechanisms within the FedAvg algorithm;
- ✓ The development of a Federated Newton Direction approach that enhances robustness against cyberattacks without transmitting gradients and Hessian matrices to the server;
- ✓ The automatic selection of fuzzy rules using genetic algorithms in expert-free environments.

Theoretical and Practical Significance

The proposed weighting approaches based on modified FIS and AHP, as well as the developed algorithms, can be utilized to enhance the performance of federated learning systems under uncertainty. These methods effectively address challenges related to data heterogeneity and network conditions in distributed environments. Furthermore, the proposed techniques support the development of adaptive federated learning systems tailored to client-specific requirements, thereby increasing system flexibility and robustness.

Approbation of the research

The research findings have been presented and discussed at international and national scientific conferences, including:

- VIII International Scientific Conference on Industrial Applications of Control and Optimization (COIA 2022), Baku, Azerbaijan;

- Republican Scientific Conference of Young Researchers and Doctoral Students (2023, 2024), Baku, Azerbaijan;
- 17th IEEE International Conference on Application of Information and Communication Technologies (AICT 2023), Baku, Azerbaijan;
- 28th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2024), Seville, Spain;
- International Conference "Artificial Intelligence: From Theory to Practice" (2024), Nakhchivan, Azerbaijan;
- ICDSM 2024, Odisha, India;
- 6th International Bosphorus Scientific Research Seminar (2025), Türkiye;
- MaCoSEP 2025 International Conference, Baku, Azerbaijan.

Publications. A total of 9 scientific works (5 journal articles and 4 conference papers) have been published on the dissertation topic. Two of the articles are indexed in Scopus-indexed journals.

Institution where the dissertation was conducted. The dissertation was conducted at the Department of General and Applied Mathematics of Azerbaijan State Oil and Industry University.

Structure and volume of the dissertation. The dissertation consists of an introduction, three chapters, a conclusion, and a list of references. The main body comprises 128 pages, including 12 figures and 8 tables. The bibliography contains 111 references. The total volume of the dissertation is 174,885 characters, distributed as follows: Introduction – 18,531; Chapter I – 49,867; Chapter II – 54,913; Chapter III – 49,515; Conclusion – 2,059 characters.

CONTENT OF THE DISSERTATION

The introduction substantiates the relevance of the research topic and the conducted studies, defines the aims and directions of the research, and presents the object and subject of the study, research methods, main provisions submitted for defense, scientific novelty obtained as a result of the research and its practical significance, the approbation of the work, as well as the volume and structure of the dissertation**.**

The first chapter is devoted to general information on Federated Learning, a novel and rapidly developing branch of machine learning. In this chapter, the existing literature is systematically analyzed, providing comprehensive information on the main directions, application domains, encountered challenges, and aggregation methods used in federated learning.

Unlike traditional centralized machine learning approaches, federated learning does not require data to be stored on a central server [1] [2].Instead, data sources—namely, clients—perform the learning process locally using their own data and transmit only the learned model parameters to the central server. One of the most significant advantages of this approach is the preservation of data privacy. Since raw data are never transferred to the central server and remain at their source locations, federated learning is particularly suitable for sensitive domains such as healthcare [3], finance, industry, and other privacy-critical sectors[4].

However, several significant challenges arise in the practical implementation of federated learning. One of the key issues is the

---

[1] Ismayilov, E., Aliyev, S., Naghiyev, E., Fatullayeva, A. Paralellization of machine learning algorithms for predicition problem using multiprocessing technique // Proceedings of the 8th International Conference on Control and Optimization with Industrial Applications (COIA-2022), – Baku: – 2022, – p. 249–251.

[2] Aliyev, S. A survey on challenges of federated learning // -Baku: Azerbaijan Journal of High Performance Computing, – 2022. 5 (2), – p. 273–285.

[3] Aliyev, S., Gozelov, T., Ceferli, T., Seyidahmedova, A. Federated learning for disease diagnosis dataset // Ümummilli lider Heydər Əliyevin anadan olmasının 100 illik yubileyinə həsr olunmuş gənc tədqiqatçı və doktorantların Respublika Elmi Konfransının Materialları, 4-cü cild. – Bakı: Azərbaycan Dövlət Neft və Sənaye Universiteti, – 2023, – p. 357–361.

[4] Drainakis, G., Katsaros, K. V., Pantazopoulos, P., Sourlas, V., Amditis, A. Federated vs. centralized machine learning under privacy-elastic users: A comparative analysis // 2020 IEEE 19th International Symposium on Network Computing and Applications (NCA), – Cambridge, MA: IEEE, – 2020, – p. 1–8.

selection and optimization of aggregation methods. Currently, the most widely used and simplest approaches are the Federated Averaging (FedAvg) and Federated Stochastic Gradient Descent (FedSGD) algorithms. These methods assign weights to client models primarily based on the number of training samples used by each client during local training. In other words, clients possessing a larger number of data samples are assigned greater weights in the aggregation process. Nevertheless, a review of the literature indicates that this approach does not always yield optimal results. In federated learning environments, clients often exhibit non-identical data distributions (non-IID settings), heterogeneous computational capabilities, and varying network conditions. If such heterogeneity factors are not properly considered, the performance of the resulting global model may degrade.

Another critical challenge in federated learning is security. Although federated learning eliminates the need for direct data transmission, the system is not entirely immune to cyberattacks. Numerous threat models have been described in the literature. For example, in gradient inversion attacks, gradients transmitted from clients to the server may be intercepted. By analyzing these gradients, an adversary may infer sensitive information about the original training data or approximate their underlying characteristics. Such attacks pose serious risks to the privacy-preserving nature of federated learning.

Therefore, alongside the development of federated learning methodologies, ongoing research focuses on improving aggregation techniques and strengthening security measures. Consequently, federated learning presents extensive research opportunities from both theoretical and practical perspectives.

The second chapter provides a more in-depth analysis of aggregation methods in federated learning and thoroughly examines their advantages and limitations.

FedAvg and FedSGD Methods in Federated Learning

One of the most widely used approaches for aggregating model parameters in federated learning is the Federated Averaging (FedAvg) algorithm. Its popularity stems from its simple structure, ease of implementation, and strong empirical performance in practical scenarios. FedAvg was proposed as an improved version of the Federated Stochastic Gradient Descent (FedSGD) method, aiming to perform optimization more efficiently in federated environments.

The fundamental idea behind FedSGD is the adaptation of the traditional Stochastic Gradient Descent (SGD) algorithm to the federated learning setting. In classical SGD, model parameters are updated using the average of gradients computed over mini-batches of data. In the federated configuration, each client computes gradients based on its local dataset, transmits these gradients to the central server, and the server aggregates (averages) them to update the global model parameters. This process can be mathematically expressed as:

$$w_{i+1} = w_i - \alpha \sum_{k=1}^{K} \frac{n_k}{n} g_k$$

Where represents the model parameters at iteration i , n_k denotes the size of the training data at client k , $g_k$is the gradient computed by client k at that iteration,

The FedAvg method was proposed to address certain limitations of FedSGD [5]. In FedAvg, clients do not transmit only gradients; instead, they update the entire set of model parameters locally for several epochs and then send the updated parameters to the server. The server aggregates these local models using weighted averaging to construct a new global model.

The update procedure can be formulated as follows:

$$Client\ k, w_{t+1}^{k} \leftarrow\ w_t^k -\ \alpha g_k$$

$$Server\ \ w_{t+1} \leftarrow \sum_{t=1}^{k} \frac{n_k}{n} w_{t+1}^{k}$$

---

[5] Gad, G., Fadlullah, Z. Federated learning via augmented knowledge distillation for heterogenous deep human activity recognition systems // Sensors, – 2023. 23 (1), – p. 1-20.

Compared to FedSGD, this approach significantly reduces communication costs and improves the efficiency of the optimization process, as multiple local updates are performed before communication with the central server.

Second order methods in Federated Learning

The Newton method is a second-order optimization technique that utilizes both the gradient and the Hessian matrix of a function to achieve faster and more accurate minimization. The core idea of the method is to approximate the objective function by a second-order Taylor polynomial and compute the optimal update step based on this approximation.

To find the local minimum of a given function $f(x)$, the Newton iteration is expressed as follows::

$$x_{i+1} = x_i - H^{-1}(x_i)\nabla f(x_i)$$

Here x_i is value of variable in i-th, $\nabla f(x_i)$ is gradient və $H^{-1}(x_i)$ is inverse of Hessian.

The Newton method can also be applied in federated learning settings and is utilized in approaches such as the Federated Newton Method (FedNM). In this framework, each client computes its local gradient and Hessian matrix and transmits them to the central server. The server aggregates this information to perform a global update.

The process involves the following steps:

- Clients compute the local gradient and Hessian matrix based on their respective datasets;
- The computed information is transmitted to the central server;
- The server aggregates the received gradients and Hessian matrices to determine the global search direction;
- The Newton iteration is applied to obtain updated global parameters;
- Clients update their local models using the new global parameters.

Although this approach provides high optimization accuracy, it introduces several significant challenges in federated environments. One of the primary concerns in FedNM is the risk to data privacy. The transmission of gradients and Hessian matrices may expose the system to attacks such as gradient inversion, enabling adversaries to infer sensitive information about the original training data.

Another major limitation relates to computational complexity. In large-scale neural networks, the Hessian matrix can be extremely high-dimensional, making its computation and storage computationally expensive. Furthermore, transmitting the Hessian matrix to the server substantially increases network traffic, leading to high communication costs in federated learning systems.

Federated Personalization

Traditional federated learning algorithms (for example, FedAvg) aim to train a single global model for all participants. However, differences in the characteristics of data stored on users' devices (such as non-uniform data distribution) may weaken the performance of the global model.

To address this issue, personalization methods have been developed. Federated Personalization (FedPer) is one such method, which creates personalized models by sharing the deep layers of the global model while keeping the final layers specific to each client.

The main idea of FedPer is that the initial layers of the model (shared layers) are common to all clients and are stored on the server as part of the global model. At the same time, the final layers (personalized layers) are kept separately for each client and are trained only using that client's local data. The reason for this is that, in neural networks, the earlier layers typically learn more general features, while the later layers learn more specific and detailed representations. Assuming that general features are similar across clients, this method allows clients to preserve the uniqueness of their models while simultaneously benefiting from the common features learned by other clients.

Thus, FedPer applies a hybrid learning approach:

- Shared deep global layers learn the overall global structure.
- Client-specific final layers are adapted to the particular data of each client.

This approach is particularly effective for systems with heterogeneous client data. The FedPer algorithm can be described as follows: Each client trains its local model, where both shared and personal layers participate. Each client sends the parameters of the shared layers to the server. On the server, the shared layers are averaged. The shared layers are then sent back to each client and combined with the personal layers.

In this algorithm, an important point to note is that the global model itself does not have decision-making capability. This is because the final output layer used for decision-making is not stored on the server and remains specific to each client. Another important note is that this algorithm cannot be applied to single-layer neural networks or logistic regression models.

Weight Assignment to Clients in Federated Learning

In federated learning, constructing a global model requires combining local models or gradients obtained from multiple clients. One of the most critical aspects of this process is determining the weight assigned to each client’s contribution. In traditional approaches, such as the FedAvg algorithm, weights are primarily determined based on the number of training samples used by each client. In other words, clients with larger datasets exert a greater influence on the formation of the global model. Although this approach is simple and intuitive, it does not always yield optimal results.

In practical federated learning environments, client data distributions may be non-uniform, class imbalance may occur, and clients may differ in computational power and network capabilities. In such cases, assigning weights solely based on the number of samples may reduce the overall quality of the global model. Therefore, recent studies propose dynamically determining weights based on additional criteria, such as class distribution, computational capacity, network latency, or data quality.

To address this issue, several methods have been employed in the research. One of these methods is the Fuzzy Inference System (FIS). The FIS was designed not only to consider the volume of data but also to incorporate client computational power and class distribution in the weighting process. Another approach is the Analytical Hierarchy Process (AHP), which was used in conjunction with the attributes applied in FIS to achieve more optimal weighting. Additionally, Genetic Algorithms and Coordinate Descent methods were utilized to iteratively improve client weights in an optimal direction at each iteration.

In the third chapter, the hyperparameters, implementation details of the proposed methods, and the obtained experimental results are discussed.

FİS

In the FedAvg method, the only factor determining the weight values assigned to clients is the number of data samples they possess. However, in certain cases, this may lead to unfair and inaccurate model training. To improve this process, we modified the FedAvg algorithm by incorporating additional factors. These factors include the level of class balance within the dataset, as well as other relevant characteristics.

These new factors were evaluated using a Fuzzy Inference System (FIS), and new weight values were calculated based on predefined rules. This evaluation enables a fairer distribution of client contributions to the federated model. In the conducted experiments, three main factors were considered: data volume, class distribution, and the client’s computational power. The membership functions are illustrated in Figures 1, 2, 3, and 4 .

After defining the membership functions, a set of rules must be established. During the experiments, the following rules were applied. These rules were used within the fuzzy inference system to determine the weight values assigned to clients [6]:

1. If Data Volume = Small and Class Balance = Highly Imbalanced, then the client’s weight is set to Weak.

---

[6] Aliyev, S., Ismayilova, N. FL2: Fuzzy logic for device selection in federated learning // IEEE Publisher: Application of Information and Communication Technologies Proceedings – 2023, – p. 1–6.

2. If Data Volume = Large and Class Balance = Balanced, then the client's weight is set to Strong.
3. If Data Volume = Small and Class Balance = Balanced, then the client's weight is set to Mid.
4. If Data Volume = Mid, then the client's weight is set to Mid.
5. If Data Volume = Large and Class Balance = Highly Imbalanced, then the client's weight is set to Mid.
6. If Computational Power (Cpower) = Bad, then the client's weight is set to Weak.
7. If Computational Power (Cpower) = Mid, then the client's weight is set to Mid.
8. If Computational Power (Cpower) = Good, then the client's weight is set to Strong.

Automatic Selection of IF–THEN Rules for FIS-Based Aggregation

One of the main challenges when using a Fuzzy Inference System (FIS) is the proper selection of rules. In expert-free environments, this process can be time-consuming and may lead to unstable results due to subjective decisions. As a solution to this problem, a Genetic Algorithm was applied. The objective was to automatically generate and optimize the rule set.

In the genetic algorithm framework, each individual represents a set of rules, and each set consists of a certain number of rules. The number of rules was dynamically adjusted and limited between two and ten. The initial rule sets were generated randomly, with the constraint that rules within the same set should not contradict one another. In each generation, the rule sets were subjected to selection, crossover, and mutation operations in order to achieve better performance. Specifically, roulette wheel selection was used in the selection phase, single-point crossover was applied during recombination, and mutation was performed by randomly modifying one selected rule.

The experiments were conducted using the Titanic dataset. The dataset columns were considered as input variables for the FIS, and corresponding membership functions were defined for each variable. The initial population started with an accuracy of approximately 33%, and after applying the genetic algorithm, the accuracy increased to 66%. This demonstrates a significant improvement in the model's learning capability. The learning curve illustrating the increase in accuracy across iterations is presented in Figure 5.

The obtained results confirm that genetic algorithms possess strong potential for the automatic selection and optimization of rules [7]. This approach can be particularly useful in environments such as federated learning systems, where expert knowledge is limited or uncertainty levels are high, thereby enhancing the effectiveness of FIS-based methods.

AHP

The Analytical Hierarchy Process (AHP) was applied as an alternative method for client weighting. The primary objective of this method is to incorporate additional factors—such as class distribution and computational power—into the weight calculation process, rather than relying solely on data quantity as in the FedAvg method. The key advantage of AHP lies in its ability to determine the relative importance of different criteria and to integrate them into the weight allocation process in a mathematically justified manner.

---

[7] Aliyev, S., Ismayilova, N, C. Zanni-Merk. Improvement of the rules selection process in FIS with genetic algorithms / -Amsterdam: Procedia Computer Science, – 2024. 246, – p. 1690–1699.

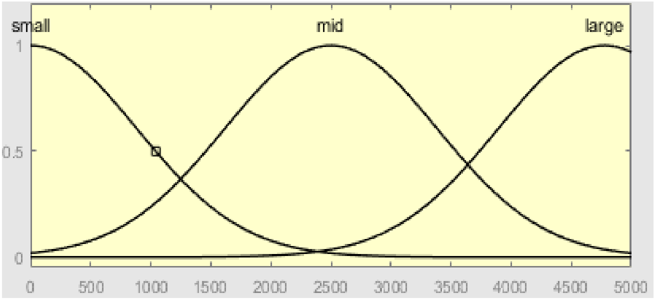

Figure 1. Membership function for the “Data Volume” variable

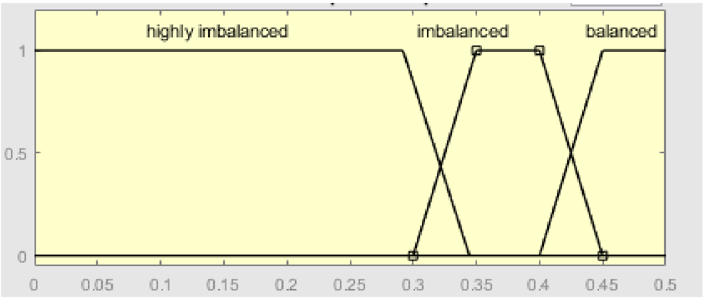

Figure 2. Membership function for the “Class balance” variable

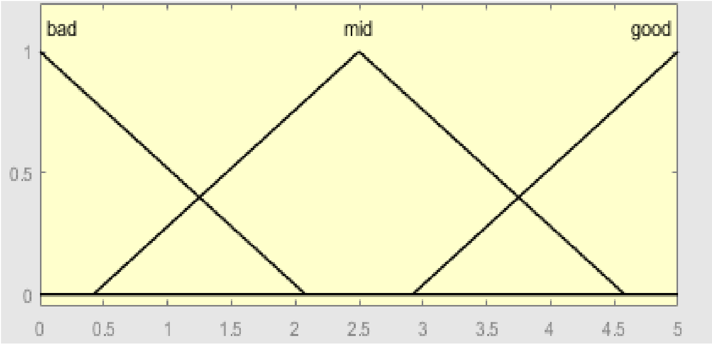

Figure 3. Membership function for the “Computation Power” variable

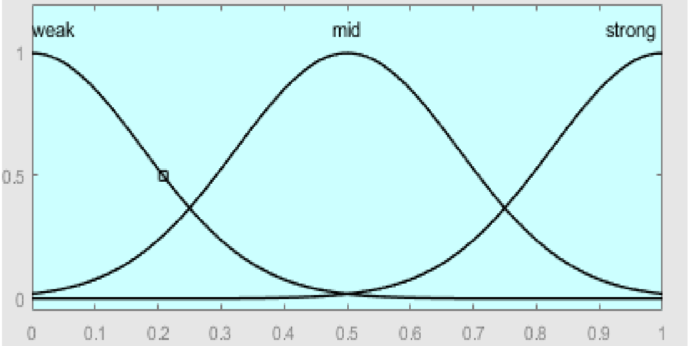

Figure 4. Membership function for the "Client weight" variable

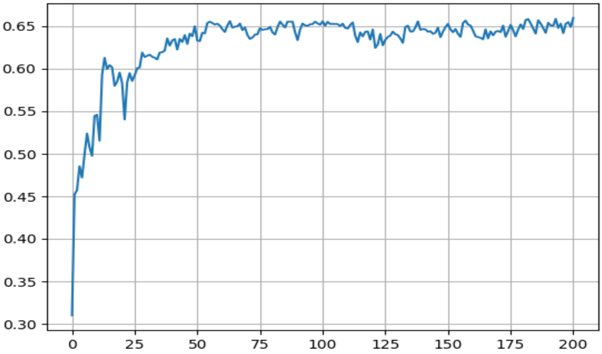

Figure 5. Learning curve of Genetic Algorithms

During the implementation of this method, three main attributes—data size, class distribution, and computational power—were considered as the principal criteria . To determine the relative importance of each criterion, a pairwise comparison matrix was constructed [8]. In accordance with the AHP methodology, this matrix was formed based on expert knowledge and logical reasoning.

---

[8] Aliyev, S. Application of AHP for weighting clients in federated learning // - Baku: Azerbaijan Journal of High Performance Computing, – 2023. Volume 6 (2), – p. 153–162.

Subsequently, the final weights of each factor were calculated using this matrix and assigned to clients as weighting coefficients.

Throughout the study, several different comparison matrices were tested, and the best-performing matrix was ultimately compared with other methods. The results indicated that class balance was the most significant factor, followed closely by data volume as the second most important criterion. Computational power was found to have comparatively lower importance than the other two criteria.

The applied comparison matrix was determined as follows:

	Size of Data	Class balance	Computation power
Size of Data	1	0.3	7
Class balance	3	1	9
Computation power	0.14	0.11	1

Coordinate descent

Coordinate Descent is one of the proposed methods, and its main distinction from other approaches is that it does not require expert knowledge or prior information about client data. This method is purely based on mathematical optimization and determines client weights by identifying the direction in which the objective function achieves an optimal value and iteratively moving toward that direction.

Initially, the weights are assigned according to the classical FedAvg method, based on data volume. Subsequently, the coordinate descent procedure is applied iteratively [9]. The process operates as follows:

- In each iteration, the weights of all clients are considered individually.
- For a given client, its weight is increased by a small coefficient α, and the updated weights are evaluated on the validation set.
- If the accuracy improves, the update continues in that direction. Otherwise, the weight is decreased by αand evaluated again.
- If neither increasing nor decreasing the weight results in performance improvement, the client's weight remains unchanged.
- After applying this procedure to all clients, the iteration is completed and a new cycle begins.
- In the next iteration, the same process is repeated starting again from the first client.
- The method continues iteratively until convergence toward the minimum of the objective function is achieved.

Experimental results

In this section, the results obtained from the above-mentioned methods are presented and comparatively analyzed. The experiments were conducted using the MAGIC Gamma Telescope dataset. For the purpose of the experiments, the dataset was manually distributed among five clients.

The distribution of the dataset across the clients is presented in Table 1. The Gini coefficient presented in the table is one of the key indicators used to describe class imbalance. These values were utilized during the weighting process in both the AHP and FIS methods. In the AHP approach, the input vector (data size, class distribution, computational power) was multiplied by the priority vector to

---

[9] Aliyev, S., Ismayilova, N. Coordinate descent optimization for device weighting in federated learning // -Baku: Proceedings of Azerbaijan High Technical Educational Institutions, – 2025. 48 (6), – p. 488–499.

calculate the final weight assigned to each client. In the FIS algorithm, these values were provided as input variables to the fuzzy inference system.

Table 2 presents the client weights obtained using the original method, as well as those calculated through the FIS and AHP algorithms.

Table 1

Distribution of MAGIC Gamma telescope dataset

Clients	Size	Class “G”	Class “H”	Gini index	Computation Power
1	4000	2000	2000	0.5	4.5
2	6300	4500	1800	0.4	3.0
3	3500	2000	1500	0.48	1.5
4	1100	500	600	0.49	4.5
5	4120	3332	788	0.31	3.0

Table 2

Weights Assigned to Clients by Different Algorithms

Clients	1	2	3	4	5
Original	0.21	0.33	0.18	0.05	0.21
FIS	0.28	0.16	0.17	0.21	0.16
AHP	0.24	0.20	0.17	0.22	0.15

Table 3

Results obtained on the MAGIC Gamma Telescope dataset.

	Accuracy				F1 Score			
Clients	Original	FİS	AHP	CD	Original	FİS	AHP	CD
1	0.75	0.77	0.78	0.77	0.74	0.78	0.76	0.77
2	0.81	0.8	0.81	0.76	0.76	0.81	0.80	0.73
3	0.78	0.79	0.79	0.78	0.76	0.79	0.78	0.78
4	0.72	0.75	0.75	0.78	0.72	0.75	0.73	0.78
5	0.84	0.83	0.82	0.78	0.72	0.82	0.83	0.70
Clinet average	0.78	0.80	0.79	0.77	0.74	0.79	0.78	0.75
Server	0.74	0.77	0.76	0.79	0.74	0.77	0.77	0.79

Table 3 evaluates the impact of different methods across various clients and presents the results in terms of both accuracy and F1-score metrics. Four different methods originally FedAvg, FIS, AHP, and Coordinate Descent (CD) were compared. The table shows the results of each method for every individual client, as well as the overall client average and the results obtained at the server level [10].

Overall, the results indicate that different methods yield different performances across clients. On average, the FIS method achieved the best results in both accuracy and F1-score metrics. Although the AHP method performed slightly behind FIS, it still demonstrated strong overall performance. The Coordinate Descent (CD) method produced good results for certain individual clients but

---

[10] Aliyev, S. Comparative analysis of different client weighting algorithms in federated learning // VII ISARC 6th International Boğaziçi Scientific Research Congress. – Istanbul: – 2025, – p. 1309–1316.

showed relatively weaker performance in terms of the overall client average. However, according to the results obtained at the server level, the CD method achieved the best performance. This suggests that although the CD method may perform weaker for individual clients, it may possess stronger generalization capability overall.

The learning curve of the Coordinate Descent algorithm is illustrated in Figure 6.

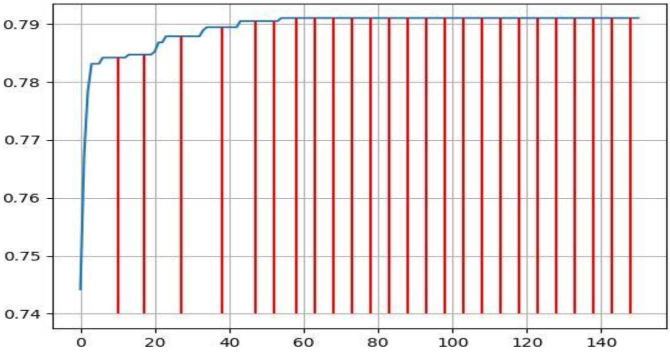

Figure 6. Learning curve of the Coordinate Descent algorithm on the MAGIC Gamma Telescope dataset.

As observed from the graph, the initial accuracy of the model was approximately 0.745. During the early stages of training, a sharp increase is observed, indicating that the model initially possesses a strong learning capability. The rapid improvement in accuracy during the first iterations confirms that the model effectively learns from the data. The red vertical lines in the graph indicate a specific point at which no changes occurred in any of the five clients; therefore, the learning parameter was reduced. The purpose of this approach is to manage convergence more precisely and ensure stability during the learning process. After reducing the parameters, the accuracy increased by approximately 0.05, resulting in a final model accuracy of 0.791.

In addition, the Coordinate Descent (CD) method was also tested on the Wisconsin Breast Cancer dataset. Since this dataset is significantly smaller than the MAGIC Gamma Telescope dataset, some of the samples stored on the server overlapped with samples replicated across clients. In this experiment, the CD method was compared with the original FedAvg algorithm.

The learning curve shown in Figure 7 confirms faster convergence of the results. Specifically, the model achieved its best performance before reaching the 20th iteration. As a result, the learning parameter was reduced only once, and high accuracy was achieved with fewer iterations. This demonstrates that the CD method performs effectively even on smaller datasets.

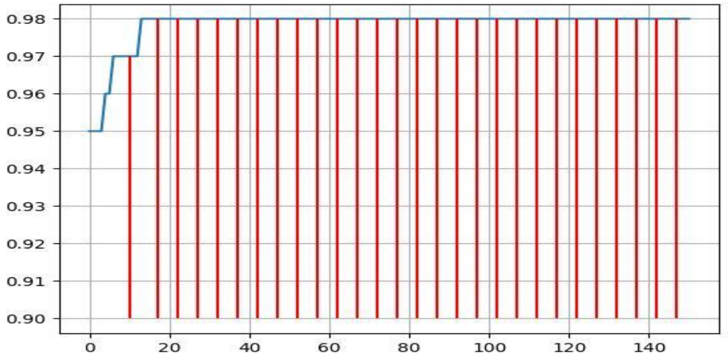

Figure 7. Learning curve of the Coordinate Descent algorithm on the Breast Cancer Wisconsin dataset.

Obtained results are demonstrated in Table 4. Here as well, the Coordinate Descent (CD) algorithm achieved an accuracy of 0.946, which is higher than the original method's result of 0.926. For the second and fifth clients, the CD method outperformed the original approach. At the server level, the CD method achieved a result of 0.98, surpassing the original method's performance of 0.95. These results indicate that the CD method generally demonstrates better overall performance and contributes to more accurate and balanced outcomes at both the client and server levels. Notably, it provides significant improvements for certain individual clients. At the server level, the higher accuracy and F1-score achieved by this method further confirm its positive impact on overall model performance.

Table 4

Results obtained on the Breast Cancer Wisconsin dataset.

	Dəqiqlik		F1 qiyməti	
Klient	Orijinal	KA	Orijinal	KA
1	0.93	0.93	0.93	0.93
2	0.91	0.97	0.97	0.97
3	0.93	0.93	0.93	0.93
4	0.94	0.94	0.94	0.94
5	0.92	0.96	0.92	0.96
Klient ortalaması	0.926	0.946	0.926	0.946
Server	0.95	0.98	0.95	0.98

FedSGD vs FedND

One of the main drawbacks of the Federated Newton method is that, at each communication round, both the gradient vector and the Hessian matrix are transmitted to the server, which increases communication overhead. Moreover, transmitting gradients to the server makes the method vulnerable to gradient inversion attacks. In such attacks, an adversary can reconstruct a substantial portion of the original data from the transmitted gradients. This, in turn, compromises the privacy preservation objective that federated learning seeks to achieve.

To address both of these issues, the Federated Newton Direction (FedND) approach was proposed. The key idea is that neither the Hessian matrix nor the gradient vector is transmitted to the server. Avoiding the transmission of the Hessian matrix reduces communication costs. Instead, each client computes its own direction vector locally:

$$S(x) = H^{-1}(x)\nabla f(x)$$

The computed direction vector is then sent to the server, where it is averaged to update the model parameters. By not transmitting the gradient directly, this method becomes more robust against gradient inversion attacks.

For weighting within this method, two different approaches were applied. First, as in the original weighting scheme, each client was assigned a weight proportional to the number of its data samples. Subsequently, for comparison purposes, weights based on the proposed FIS approach were tested. Both weighting strategies were applied to the FedSGD and FedND methods. The experiments were conducted using the MAGIC Gamma Telescope dataset.

Within the framework of comparing FedSGD and FedND algorithms, various aspects of both methods were thoroughly analyzed. The primary objective of this comparison was to evaluate their overall performance and effectiveness, as well as to investigate the characteristics of the applied weighting mechanisms. Accuracy, AUC, and F1-score were selected as evaluation metrics, enabling an objective assessment of the algorithms across multiple dimensions.

The obtained results indicate that the FedND algorithm outperformed FedSGD across all evaluation metrics and produced more stable results for each client. This superiority may be attributed to its more effective optimization direction. On the other hand, when FIS-based weighting was applied, its impact varied across different metrics. For the FedND algorithm, FIS provided additional improvement mainly in terms of accuracy, while in other metrics it produced results close to those obtained with the original weighting scheme. In contrast, for FedSGD, the original weighting generally underperformed compared to FIS-based weighting, except for a slight advantage observed in the F1-score.

Table 5 presents the results obtained for both algorithms using original and FIS-based weighting schemes. Additionally, it specifies which configuration achieved the best performance for each client, thereby enabling comparison at both global and local levels.

Table 5

Results of the FedSGD and FedND algorithms. The bolded values indicate the best-performing results.

Algorithm	Criterion	Client 1	Client 2	Client 3	Client 4	Client 5	Average
FEDSGD Original	Accuracy	0.70	0.77	0.69	0.68	0.81	0.73
	AUC score	0.717	0.714	0.679	0.713	0.684	0.702
	F1 score	0.71	0.77	0.70	0.69	0.80	0.734
FedSGD with FIS	Accuracy	0.71	0.77	0.71	0.70	0.79	0.736
	AUC score	0.715	0.713	0.698	0.719	0.678	0.704
	F1 score	0.71	0.76	0.71	0.71	0.77	0.732
FedND Original	Accuracy	0.74	0.81	0.78	0.72	**0.84**	0.778
	AUC score	0.763	**0.776**	**0.781**	**0.762**	**0.733**	**0.763**
	F1 score	0.75	**0.82**	0.78	0.73	**0.84**	**0.784**
FedND with FIS	Accuracy	**0.76**	**0.81**	**0.78**	**0.74**	0.82	**0.782**
	AUC score	**0.763**	0.762	0.768	0.761	0.705	0.751
	F1 score	**0.76**	0.81	**0.78**	**0.74**	0.81	0.78

One of the main disadvantages of the FedND algorithm is the need to compute the inverse of the Hessian matrix at each iteration. Although the Newton method converges more rapidly, calculating the matrix inverse increases the computational complexity of the algorithm. In this study, the Hessian matrix was not computed explicitly; instead, the direction vector was obtained by solving a system of linear equations. This approach resulted in relatively faster execution of the algorithm.

Thus, by using the FedND algorithm, we achieve both faster convergence and improved robustness against gradient inversion attacks by eliminating the direct transmission and use of gradients.

FedPER

Table 6 below presents the results obtained from applying the FIS and AHP algorithms to the FedPER algorithm

Table 6

Results of the FedSGD and FedND algorithms.

Method	Accuracy	Precision	Recall	F1 Score
Original	0.4125	0.4724	0.4125	0.4412
FIS	0.4325	0.5118	0.4188	0.4616
AHP	0.4325	0.5092	0.4188	0.4613

These results were obtained using the CIFAR-10 dataset. As in the previous experiments, several different weighting criteria were tested and the obtained results were compared. In the experiment, the initial layers of the convolutional neural network were shared among all clients, while the last three layers were kept personalized. The model's performance was evaluated on each client's test set, and the results were assessed based on their average values [11].

In the experimental phase, the Convolutional Neural Network (CNN) architecture was optimized according to the principles of personalized federated learning. Within this framework, the initial layers of the network (feature extraction layers) were shared among all clients as a common knowledge base, while the final three layers were kept private to adapt to the local data characteristics of each client. This hybrid structure facilitates both the acquisition of global

---

[11] Aliyev, S., Dadasov, A., Nabiyeva, F. Comparative analysis of different approaches for aggregation step in federated fearning // International Conference on Artificial Intelligence: from theory to practice, – Nakhchivan: – 2024. – p. 278–284.

knowledge and local adaptation. Model performance was verified on each client's specific test sets, and the results were evaluated based on statistical averages. The analysis indicates that the original FedAVG weighting performed significantly worse than the other innovative approaches across all four evaluation criteria. Notably, the FIS-based approach outperformed AHP-based weighting in all scenarios due to its superior management of uncertainty. These empirical findings confirm the crucial role of intelligent weighting methods in personalized federated learning and open broad scientific perspectives for their further development.

The obtained results indicate that, across all four evaluation metrics, the original weighting scheme performed weaker than the other approaches. The FIS-based weighting consistently outperformed the AHP-based weighting in all cases. This finding demonstrates the effectiveness of methods such as FIS and AHP, particularly in personalized federated learning environments, and highlights the significant potential for their further development.

MAIN SCIENTIFIC RESULTS

As a result of the theoretical and practical research conducted within the framework of the dissertation, the following important scientific results were obtained:

1. Conceptual advantages of federated learning and an analysis of the weighting problem were carried out. Ensuring data privacy in modern distributed systems is a significant priority. The research has shown that, unlike centralized approaches, federated learning enables model training on local devices without the need to transmit raw data, thereby protecting user privacy at a fundamental level. However, the internal mechanism of the traditional FedAVG algorithm does not account for the quality differences of local devices, as it relies solely on the quantity of data. The analyses conducted prove that excluding factors such as class balance, computational resources of the devices, and training frequency results in a decline in the overall performance of the model. To address this deficiency, new weighting strategies based

on the principles of both fairness and computational efficiency have been proposed in this work.

2. The integration of multi-criteria decision-making and optimization methods was implemented. To solve the aforementioned structural problems, methods with different characteristics, such as FIS, AHP, coordinate descent, and genetic algorithms, were applied in the dissertation. The FIS and AHP approaches transform expert opinions into a mathematical model, integrating data quality and the technical power of the device into a single weighting coefficient along with the volume of data. On the other hand, optimization-based approaches, namely coordinate descent and genetic algorithms, have minimized the system's dependence on human intervention by ensuring the automatic and dynamic adjustment of weights. Tests conducted on both FedAVG and personalized FedPER models demonstrated that the proposed methods significantly outperform standard methods in key metrics such as accuracy, F1-score, and AUC.

3. The automation of expert knowledge and the automatic selection of the rule base were achieved. One of the greatest challenges encountered in the construction of expert systems is the elimination of subjectivity and the formation of the rule base. To solve this problem, a mechanism for the automatic selection of rules was developed in the study by utilizing the capabilities of genetic algorithms. Experiments conducted on the Titanic dataset confirmed the effectiveness of this approach: the system, which showed only 33% accuracy with random rules in the initial stage, evolved through the application of genetic operators and increased its accuracy rate to 66%. This result proves that genetic algorithms play a decisive role in enhancing the self-learning and adaptation capabilities of fuzzy systems.

4.Second-order optimization was applied, and the level of cyber-resilience was increased. The final important direction of the research is related to increasing the convergence speed and security of the model in a federated environment. The application of the Newton direction method, a second-order optimization technique, has shown that this method allows for reaching a higher level of accuracy with fewer iterations compared to standard gradient-based methods.

Additionally, due to this method's sensitivity to data structure, it was determined to exhibit higher resilience against cyber-attack scenarios such as model poisoning and gradient inversion. Thus, the proposed approach both increases the learning speed of the model and makes it more reliable against external interferences.

The main content dissertation of the dissertation is published in the following works:

1. Ismayilov, E., Aliyev, S., Naghiyev, E., Fatullayeva, A. Paralellization of machine learning algorithms for predicition problem using multiprocessing technique // Proceedings of the 8th International Conference on Control and Optimization with Industrial Applications (COIA-2022), – Baku: – 2022, – p. 249–251.
2. Aliyev, S. A survey on challenges of federated learning // - Baku: Azerbaijan Journal of High Performance Computing, – 2022. 5 (2), – p. 273–285.
3. Aliyev, S., Gozelov, T., Ceferli, T., Seyidahmedova, A. Federated learning for disease diagnosis dataset // Ümummilli lider Heydər Əliyevin anadan olmasının 100 illik yubileyinə həsr olunmuş gənc tədqiqatçı və doktorantların Respublika Elmi Konfransının Materialları, 4-cü cild. – Bakı: Azərbaycan Dövlət Neft və Sənaye Universiteti, – 2023, – p. 357–361.
4. Aliyev, S., Ismayilova, N. FL2: Fuzzy logic for device selection in federated learning // IEEE Publisher: Application of Information and Communication Technologies, Proceedings. – 2023, – p. 1–6. **(Scopus, WoS).**
5. Aliyev, S. Application of AHP for weighting clients in federated learning //-Baku: Azerbaijan Journal of High Performance Computing, – 2023 6 (2), – p. 153–162.
6. Aliyev, S., Ismayilova, N., Zanni-Merk, C. Improvement of the rules selection process in FIS with genetic algorithms / - Amsterdam: Procedia Computer Science, – 2024. 246, – p. 1690–1699. **(Scopus).**

7. Aliyev, S., Dadasov, A., Nabiyeva, F. Comparative analysis of different approaches for aggregation step in federated learning //International Conference on Artificial Intelligence: from theory to practice, –Nakhchivan: – 2024. – p. 278–284.
8. Aliyev, S. Comparative analysis of different client weighting algorithms in federated learning // VII ISARC 6th International Boğaziçi Scientific Research Congress, – Istanbul: – 2025, – p. 1309–1316.
9. Aliyev, S., Ismayilova, N. Coordinate descent optimization for device weighting in federated learning // -Baku: Proceedings of Azerbaijan High Technical Educational Institutions, – 2025. 48 (6), – p. 488–499.

Author's personal role in co-published works:

[1] – Computer simulation and analysis of results;
[3] – Conceptualization and computer simulation;
[4] – Computer simulation, data preprocessing, and analysis of results;
[6] – Computer simulation and analysis of results;
[7] – Problem statement, computer simulation, and analysis of results;
[9] – Conceptualization, data preprocessing, computer simulation, and analysis of results.

The defense will be held on 28 April 2026 at 16:00 at the meeting of the Dissertation council FD 2.48 of Supreme Attestation Commission under the President of the Republic of Azerbaijan operating at Azerbaijan State Oil and Industry University.

Address: AZ1010, Baku, Azadlıq avenue 16/21

Dissertation is available at the library of Azerbaijan State Oil and Industry University.

Electronic versions of the dissertation and abstract are available on the official website of Azerbaijan State Oil and Industry University.

The abstract was sent to the necessary addresses on 17.03.2026.

Signed to print. 11.03.2026
Paper format: A5
Volume: 36345
Number of hard copies: 20