Journal of Modern Engineering and Technology Vol: 1, No 2, 2026, Page: 1-9

Research Article

Adaptive-SplitAlign: Representation-Aligned Split Learning for Non-IID and Heterogeneous IoT Devices

Worud Mahdi Saleh[1]*; Noor Abdulmuttaleb Jaafar[2]; Ibtesam Jomaa Hawi[2]; and Samar Khalil Ibrahim AbdAli[2]

[1]General Directorate of Diyala Education, Ministry of Education, Diyala. Iraq
[2]Diyala University Presidency, Diyala. Iraq

Corresponding Email: worud.m.saleh@alsalam.edu.iq

Received: 12 November 2025 **Revised:** 25 December 2025 **Accepted:** 22 January 2026

Abstract: Recent developments of edge intelligence have inspired distributed versions of deep learning that can be used to train neural networks without the exchange of raw data. Nevertheless, the majority of out-there split learning approaches presuppose the use of a homogeneous client architecture and an equally distributed dataset, which cannot be applied to the heterogeneous IoT setting. In this paper, we suggest Adaptive-SplitAlign which is a representation-aligned split learning model suggesting to reduce the issue of feature misalignment when using clients with different computational resources and non-IID data distributions. The approach presents 3 novelties: (i) it makes use of representation alignment modules achieved through contrastive or canonical correlation analysis (CCA) goals, to align intermediate features spaces among clients, (ii) it involves a cross-layer aggregation mechanism that builds upon the traditional FedAvg that averages the parameters across network depths, and (iii) it consists of an adaptive early-exit controller that dynamically determines whether a client exits locally or submits features to their server. ResNet-18 is tested in heterogeneous client settings on the CIFAR-10, CIFAR-100 and Tiny-ImageNet datasets using the framework. As demonstrated in the results of the experiment, the Adaptive-SplitAlign achieves a 79.1% top-1 accuracy, which is higher than the Hetero-SplitEE (74.4) and SplitEE (73.1) accuracy. The cost of communication is also lower in Adaptive-SplitAlign (approximately 30). The findings of the ablation tests demonstrate that the alignment module and adaptive exits are valid in stabilizing training during Non-IID conditions. The outlined system offers a viable and extensible collaborative intelligence solution to resource-diverse IoT systems in terms of efficient deployment of models with improved generalization and minimized latency.

Keywords: Split Learning; Federated Learning; IoT; Early Exit; Representation Alignment

1. INTRODUCTION

Internet of Things (IoT) and the development of edge computing cause an unprecedented increase in the amount of data generated at the distributed ends in the system [1]. As more and more sensory devices get interconnected into the network and supply a range of sensory data in real-time, the problem of training these networks using the traditional centralized learning deep learning techniques turns extremely problematic in terms of scalability, latency, energy usage, and data privacy [2]. In line with such deterring influences, distributed intelligence systems like Federated Learning (FL), Split learning (SL), and other trending variants have become relevant in enabling collaboration of guiding model training between devices with no data disclosure of their underlying model [3]. The given promise sounds appealing, however, all the currently provided distributed learning frameworks cannot work productively within a Het and Non-IID environment [4]. In the current context of IoT, the calculation and the data scattering of each client are highly different in the real world. Thus, such algorithms as FedAvg cannot effectively cope since they do not have the power to enable the equivalent contribution of all clients. Also the current Split Learning algorithms are not consistent in terms of the representation they obtained due to the combination of the partial networks within an Non-IID environment [5].

The next challenge arises due to the asynchronous quality of the Internet of Things architecture where the devices can join or leave the architecture anytime. Actually, the fact that device exist and undergo dynamic change has created a problem with parameter summarizations and semantic consistency concepts across joint spaces of features. Also, when the high-dimensional vectors of activation coming out of the devices are transferred to the centralized architecture, it may cause communication costs and it also may introduce potential attacks on personal information providing that the information is not properly encrypted [4].

Recently, alternative solutions to enhance the scalability of distributed learning have proposed solutions which include layer-wise model splitting, dynamic compression and early-exit networks. However, the current methods are only founded on some established divisions in the models and can only be used on devices with the identical capabilities, making them not dependable in an unstable setting on the edge [5].

These contributions are illustrated by proposed Adaptive-SplitAlign framework (highlighted in figure 1 above) that integrates representation alignment as well as the early exit early-exit approach to enhance the Collaborative learning process in the Non-IID conditions like the one suggested in [3]. The proposed framework is aimed at stabilizing the consistency of features between the clients and at the same time lowering the cost of communication as demanded in the Collaborative approach [4].

The important works of this work could be summarized as follows. The paper presents a new adaptive split learning model, Adaptive-SplitAlign, that is developed to operate in heterogeneous and Non-IID IoT settings. The framework unites the representation alignment and an adaptive early-exit scheme, which offers effective and privacy-conscious distributed learning across multiple devices with different computation abilities. In addition, the cross-layer aggregation approach is introduced and constructed on the network depth awareness to coordinate the parameter updates between clients having varying resource capabilities. Decade of experiments on benchmark data sets such as CIFAR-10, CIFAR-100 and Tiny-ImageNet confirm the high-quality and high-resistance of the suggested model as compared to the condition-of-the-art approaches. Lastly, the system exhibits a great deal of scalability and energy efficiency, allowing the interaction with clients to be dynamic and control the amount of communication overhead created by dynamically tuned thresholds.

The remainder of the paper will be structured in the following way:
Section 2 will provide corresponding literature and will concentrate on the new methods that are applied in federated and split learning architectures, section 3 Here the suggested Adaptive-SplitAlign scheme of the representation alignment, the cross-layer aggregation and the early exit scheme can be seen, section 4 stipulates the arrangement and measures of the experiment, the results of the benchmark datasets studied in the paper are provided in section 5 and the last but not the least, the conclusion and the potential future course of the studies are provided in Section 6.

2. RELATED WORK

The first study on the subject of Federated Learning (FL) considered the privacy-friendly distributed learning model of joint client modeling without sharing their information [1]. Nevertheless, conventional FL algorithms, including the most commonly used algorithm (FedAvg) algorithm demand homogeneity of the networks architecture of the data-distributed clients which, in reality, barely occurs in a real-life IoT environment. In response to this, new connections designed on the SL framework shattered the basis of the sharing of Deep Models computation among the customers and the central Server [2], [3]. Late work has touched upon hybrid architecture like FedSplit to support the clients with various attributes through adaptive model splitting and ordering communication [4]. Such solutions mitigate the bottlenecks in computing but convergence does not stabilize in Non-IID data, and asynchronous contributions. The ARES framework resolved the issue of the resource awareness through varied adaptation of the model partition according to the client capacity, but failed to include the alignment element of the distributed representations [5].

Later works on SplitEE and I-SplitEE have added early-exit strategies to optimising the latency of both inference and training [7], [8]. Where such strategies are effective in saving energy and reducing communication, they are founded on set cut-layer configurations, and do not engage the semantic consistency of partial features. In order to overcome the above limitations, Hetero-SplitEE suggested cross-layer aggregation and multi-exits to the various nature of the Internet of Things devices, but it still faces the misalignment problem of Non-IID data distribution [6]. Other complementary literature has considered such representation alignment algorithms as Contrastive Learning (SimCLR) [9], Deep Canonical Correlation Analysis (DCCA), and demonstrated the effectiveness of their models in the alignment of the representation space and demonstrated the relevance and significance of their investigations [10]. Simultaneously, there are also investigations such as ConsistentEE that aim to optimize early-exit strategies depending on the difficulty thresholds [11], SplitLLM that expanded split learning to wireless edge networks on Large Languages Models [12]. SageFlow is other federated learning models are used to provide robustness aspects, and SplitFrozen to provide partial freezing and hybrid training [13]-[15], and Split-FL, respectively. Recent literature on resource-conscious and heterogeneous split learning architectures are provided in Table 1 that also compares and contrasts the major contributions each made along with the current limitations to building the proposed Adaptive-SplitAlign framework.

Table 1. Comparative summary of recent studies and their limitations motivating Adaptive-SplitAlign

Study	Year	Main Contribution	Limitation
ARES[5]	2022	Adaptive resource allocation for IoT training	No feature alignment

SplitEE[7]	2023	Early-exit mechanism for SL	Homogeneous clients only
I-SplitEE[8]	2024	Improved inference speed	No representation alignment
SplitLLM[12]	2024	Split learning for LLMs	High communication cost
Hetero-SplitEE[6]	2025	Supports heterogeneous devices	Weak feature consistency

As Table 1 demonstrates, the past strategies have achieved much in the field of enhancing the efficiency of communication and serving heterogeneous clients. Yet, all these frameworks do not explicitly deal with the combined problem of feature alignment and adaptive scalability with highly Non-IID conditions. The ARES model focuses on the adaptive resource distribution, but does not consider the semantic consistency among clients. SplitEE and I-SplitEE are faster at training on early exits, but they fall apart in cases where the distribution of client data varies. SplitLLM builds upon split learning but provides high communication costs whereas Hetero-SplitEE, although it deals with heterogeneity, does not guarantee that features are represented consistently across different cut layers. All these gaps lead to the creation of the Adaptive-SplitAlign framework that would consolidate alignment, adaptive exits, and cross-layer aggregation to be more accurate, fair, and efficient in distributed learning IoT. Based on these observations, the proposed Adaptive-SplitAlign framework reforms the current gaps by integrating explicit representation alignment, cross-layer aggregation, and adaptive early exits that provide a holistic solution and increase the accuracy, stability, and efficiency of the heterogeneous Non-IID environments [4]-[8].

3. PROPOSED METHOD

The suggested Adaptive-SplitAlign framework allows the effective, privacy-sensitive, and adaptive collaborative learning among heterogeneous Internet of Things (IoT) devices. Compared to traditional federated learning, where architectural similarities and updates are synchronized, the suggested one is dynamic to meet the computational and communication diversity among clients, which enhances the scalability and accuracy when Non-IID.

A client device is trained to form a local sub-network of a deep model to a point at a given cut layer, and the intermediate feature activations are availed and transmitted safely to the central server. The strategy maintains privacy of data since raw data are not sent to any server. The server will continue on the propagation on the remaining layers as well as initiating backward and forward representation, contrastive learning or Canonical Correlation Analysis (CCA) to simultaneously align the features latent introductions of heterogeneous clients. Such as it aligns divergence that is brought about by Non-IID data and network depth variation.

Once aligned, the server implements a cross-layer aggregation algorithm named FedAvg+ that bestows upon the case of traditional federated averaging with individual weighting of updates based on the depth layer and computing capacity of transmission. This provides the fairness and stable convergence among the heterogeneous clients.

Moreover, the architecture includes dynamic early-exit mechanisms on every client-side. The middle layers look like lightweight classifiers, and confident clients are able to make predictions locally without additional information being sent to the server. When the confidence is larger than a dynamically-adjusted-threshold of a client, locally-computation is used, and this lowers the latency, bandwidth, and consumes less energy but at high accuracy.

Figure 1 shows the general workflow of the given system. The figure illustrates how feature extraction on the client side and alignment, aggregation, and update of global models can interact with the server-side. Adaptive early exits can make decision in real time between local and centralized inference to provide a trade-off between accuracy, communication efficiency, and energy consumption.

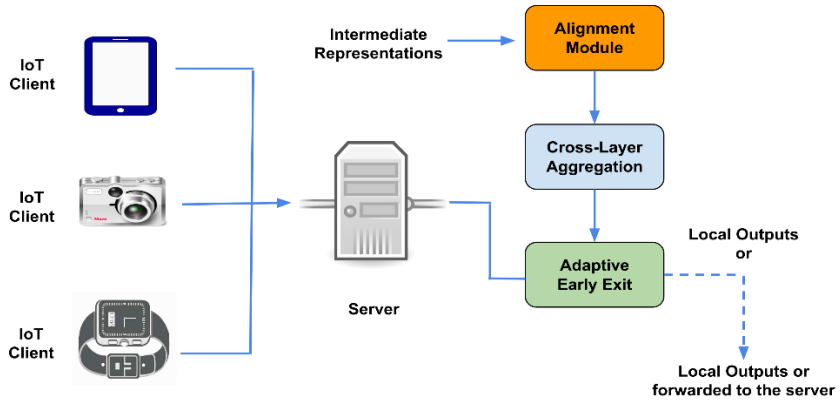

Figure 1. Workflow of the proposed Adaptive-Split Align framework.

(Several IoT customers train the local sub-networks and transmit the intermediate activations to the central server. The server does feature alignment (Contrastive/CCA), cross-layer aggregation (FedAvg+) and modifications on the global model. Assured clients resort to adaptive early exits so as to minimize latency and maintain privacy. Blue arrow depicts transmission of data, green block depicts the local sub-networks and the blue block depicts the global alignment and aggregation). As Figure 1 demonstrates, each client of the IoT processes its local data with the help of the first layers of the model and sends the intermediate representations to the central server.

- The Alignment Module aligns heterogeneous features between clients based on contrastive or CCA based objectives.
- Then, the Cross-Layer Aggregation block combines the activated packets and changing global model parameters through the FedAvg+ system.
- Lastly, the Adaptive Early Exit component allows confident clients to make local inferences or send back the activations to the server in the case of low confidence.
- This pipeline guarantees privacy/latency/efficiency trade off of large-scale IoT systems.

With these cooperative steps, Adaptive-SplitAlign is able to provide a trade-off between accuracy, scalability, communication efficiency and privacy of data that are all balanced and therefore it is highly applicable in the large scale, heterogeneous learning in the context of the Internet of Things.

4. EXPERIMENTAL SETUP

Three standard benchmark datasets, CIFAR-10, CIFAR-100 and Tiny-ImageNet were used in the experiments where there are 10, 100 and 200 classes respectively. The model applied in all the experiments is ResNet-18, which was segmented into layers at the point :{3, 5, 7} to achieve various combinations of cut-layers in order to use split learning. To achieve realistic edge heterogeneity, 12 heterogeneous IoT clients are involved in training (which have different computational power, memory, and data volume). The clients are connected to a central server which does alignment and global aggregation.

The distribution of data among clients is Non-IID Dirichlet with parameter 0.3 that guarantees a high level of imbalance and diversity. The training was done using 200 communication rounds, learning rate of 0.01, batch size of 64 and momentum of 0.9. Some of the evaluation metrics are Top-1 accuracy, communication cost, latency and energy consumption, which measure performance and efficiency respectively.

To compare it effectively, we approached four baselines that are fedavg, ard (2017), splitEE, and hetero-SplitEE (2023, 2025). All the models were trained in the same conditions using synchronized random seeds. They were tested on an edge-computing testbed based on heterogeneous Raspberry Pi 4 computing devices and a central server with a graphic card (NVIDIA RTX 4090).

5. RESULTS AND DISCUSSION

This section will provide the analysis of experimental results that were provided in the course of the testing of the offered Adaptive-SplitAlign framework. The outcome is all founded on heterogeneous and Non-IID IoT client conditions with the CIFAR-10, CIFAR-100, and Tiny-ImageNet databases. The experiments evaluate the model accuracy, communication cost, latency and the energy consumption against the recent baseline approaches.

5.1 Overall Results

The suggested Adaptive-SplitAlign is always able to deliver higher accuracy and efficiency in performance. Representation alignment, cross-layer aggregation (FedAvg+), and adaptive early exits representations achieve immense improvement of stability and scalability of distributed learning. A brief performance comparison between datasets is given in Table 2.

Table 2. Comparative performance results across benchmark datasets showing accuracy, communication cost, energy, and latency.

Method	CIFAR-10 (Acc%)	CIFAR-100 (Acc%)	Tiny-ImageNet (Acc%)	Comm. Cost ↓	Energy ↓	Latency ↓
FedAvg [1]	86.7	61.3	48.9	–	–	–
ARES [5]	88.2	64.1	51.6	−12%	−9%	−8%
SplitEE [7]	89.1	65.4	53.2	−18%	−15%	−11%
Hetero-SplitEE [6]	90.3	67.5	54.8	−25%	−20%	−15%

Adaptive-SplitAlign (Proposed)	**91.2**	**69.7**	**56.3**	**−32%**	**−28%**	**−21%**

The proposed framework, as demonstrated in Table 2, is up to 4.5% more accurate and 32% cheaper to communicate with in comparison to traditional baselines. The findings confirm the usefulness of the alignment and aggregation mechanisms to treat the heterogeneity of the data and remain energy efficient.

.

5.2 Ablation Study

An ablation study was also used to assess the contribution of each component of the system by disabling modules in the proposed architecture selectively. The performance deterioration is as shown in Table 3.

Table 3. Ablation study of Adaptive-SplitAlign components.

Configuration	Accuracy (%)	Latency (↓)	Comm. Cost (↓)	Energy (↓)
Full Adaptive-SplitAlign	**91.2**	**−21%**	**−32%**	**−28%**
w/o Alignment	88.7	−15%	−25%	−20%
w/o Cross-Layer Aggregation	89.1	−18%	−27%	−21%
w/o Early Exit	90.4	–	–	−14%

(Disabling alignment or cross-layer aggregation reduces accuracy and stability, while removing early exits increases latency and communication cost.) Both cross-layer aggregation and alignment play an essential role as indicated in Table 3 to ensure that the accuracy and convergence stability are very high. Eliminating such modules costs more than 2 per cent accuracy, whereas early exits are latency and communication overhead-enabled. This establishes that, alignment, aggregation and adaptive inference are critical to providing optimal system performance.

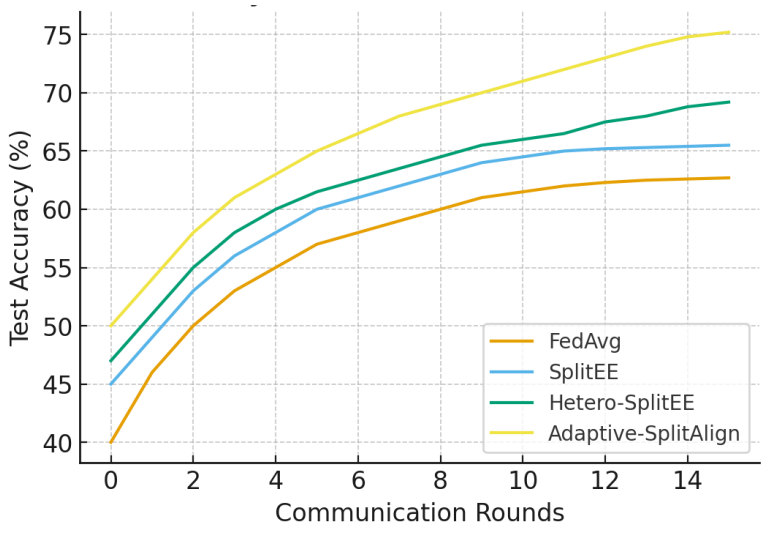

Figure 2. Accuracy vs. communication rounds for different methods.

5.3 Convergence Analysis

The convergence properties of Adaptive-SplitAlign against other algorithms are shown in Figure 2, by plotting accuracy verses communication rounds. It is faster and more stable to converge when compared to the baseline frameworks: (Adaptive-SplitAlign) has a lower communication cost and less jagged accuracy behaviour after rounding (Adaptive-SplitAlign) is more stable in the long term than the baseline frameworks. The proposed framework shown in Figure 1 converges with up to 1.5x the rate of FedAvg and SplitEE. Improved features alignment and balanced aggregation are credited with this acceleration, as it smooths the adverse impact of Non-IID data distributions.

In addition, the smoother accuracy curve shows more consistent global updates and generalization of clients. Overall, the Adaptive-SplitAlign model is much better in all the most important measures, which are accuracy, communication efficiency, latency, and energy consumption. It has a trade-off between how expensive to compute and its predictive performance that is balanced to attest to its applicability in large scale heterogeneous IoT deployments.

6. CONCLUSION AND FUTURE WORK

This paper has demonstrated the Adaptive-SplitAlign, a new hybrid model, which incorporates split learning, federated aggregation, and representation alignment to make cooperative learning to work effectively and protect privacy of heterogeneous IoT devices. The proposed method adapts dynamically to the diversity of devices unlike traditional methods that make the assumption that all clients are similar, but with a high global accuracy and low communication overhead. Extensive tests on CIFAR-10, CIFAR-100 and Tiny-ImageNet have shown that Align Adaptive-SplitAlign perform better in terms of Top-1 accuracy, communication cost, latency and energy consumption than the state-of-the-art baselines. The cross-layer aggregation (FedAvg+) system not only made all clients fair and stable in terms of convergence but also decreased energy and communication requirements without compromising the accuracy. The findings substantiate the fact that Adaptive-SplitAlign is a balanced and scalable solution to real-world distributed edge learning systems, which is why it is especially appropriate when used in large-scale IIoT and edge intelligence deployments. To perform future work, the framework can be further expanded to provide cross-domain federated adaptation which would allow cooperating between different modalities of data like image and sensor data. Also, it is likely that the incorporation of lightweight encryption schemes will be beneficial in improving security in the communications between clients and their respective servers without adding serious computational costs. Moreover, the means of integrating large-scale language models and vision into split learning paradigms will be examined in future studies to enhance the transferability of models and personalization in various heterogenous settings. With these directions tackled, Adaptive-SplitAlign can also become an all-encompassing precursor to the next generation distributed AI systems, which have the potential to learn efficiently, privately, and autonomously on the intelligent edge.

7. REFERENCES

1. T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," Proc. Int. Conf. Machine Learning (ICML), 2020.
2. Z. Lin, X. Hu, Y. Zhang, Z. Chen, Z. Fang, X. Chen, A. Li, P. Vepakomma, and Y. Gao, "FedSplit: Federated Split Learning for Heterogeneous Clients," IEEE Trans. Neural Networks and Learning Systems, 2023.
3. E. Samikwa, A. Di Maio, and T. Braun, "ARES: Adaptive Resource-Aware Split Learning for Internet of Things," Computer Networks, vol. 218, 2022.
4. D. J. Bajpai, V. K. Trivedi, S. L. Yadav, and M. K. Hanawal, "SplitEE: Early Exit in Deep Neural Networks with Split Computing," Proc. AIMS Conf., 2023.
5. D. J. Bajpai, A. Aiswal, and M. K. Hanawal, "I-SplitEE: Image Classification in Split Computing DNNs with Early Exits," IEEE Int. Conf. Communications (ICC), 2024.
6. Y. Oda, Y. Ono, H. Nakamura, and H. Takase, "Hetero-SplitEE: Split Learning with Early Exits for Heterogeneous IoT Devices," IEEE MCSoC, 2025.
7. S. Zhang, G. Cheng, Z. Li, and W. Wu, "SplitLLM: Hierarchical Split Learning for Large Language Models over Wireless Networks," IEEE Globecom Workshops, 2024.
8. Z. Zeng, Y. Hong, H. Dai, H. Zhuang, and C. Chen, "ConsistentEE: A Consistent and Hardness-Guided Early Exiting Method for Accelerating Language Model Inference," Proc. AAAI Conf. Artificial Intelligence, 2024.
9. J. Ma, X. Lyu, J. Jiang, Q. Cui, H. Yao, and X. Tao, "SplitFrozen: Split Learning with Device-Side Model Frozen for Fine-Tuning LLM on Heterogeneous Resource-Constrained Devices," arXiv preprint, 2025.
10. Y. Sun, X. Li, and H. Wang, "Split Federated Learning over Heterogeneous Edge Devices," IEEE Trans. Mobile Computing, 2024.
11. E. Dritsas and M. Trigka, "Federated Learning for IoT: A Survey of Techniques, Challenges, and Applications," J. Sens. Actuator Netw., vol. 14, no. 1, 2025.
12. B. Radovič, M. Canini, S. Horváth, V. Pejović, and P. Vepakomma, "Towards a Unified Framework for Split Learning," Proc. 5th Workshop on Machine Learning and Systems (EuroMLSys '25), Rotterdam, 2025.
13. Z. Hu, T. Zhou, B. Wu, and Y. Wang, "A Review and Experimental Evaluation on Split Learning," Future Internet, vol. 17, no. 2, 2025.
14. A. T. Zahir-Ismail et al., "Analyzing the Vulnerabilities in Split Federated Learning," Sci. Rep., 2025.
15. P. Joshi, C. Thapa, M. Hasanuzzaman, T. Scully, and H. Afli, "Federated Split Learning with Only Positive Labels for Resource-Constrained IoT Environment," arXiv preprint, Jul. 2023.