

A Federated Learning-Based Virtual Interpreter for Arabic Sign Language Recognition in Smart Cities

Ahmad Alzu'bi¹, Tawfik Al-Hadhrami², Amjad Albashayreh³, Lojin Bani Younis¹

¹*Department of Computer Science, Jordan University of Science and Technology, Irbid, Jordan*
agalzubi@just.edu.jo, lhbaniyounis19@cit.just.edu.jo

²*School of Science and Technology, Nottingham Trent University, Nottingham, UK*
tawfik.al-hadhrami@ntu.ac.uk

³*Department of Computer Science, The University of Jordan, Amman, Jordan*
amalbashayreh20@cit.just.edu.jo

Abstract- Arabic Sign Language (ASL) exhibits structured grammar and syntax, necessitating adherence to these rules in automated sign language generation for people with disabilities. Hence, the effective generation of sign language relies heavily on using 3D virtual signers. This paper presents a groundbreaking framework that employs federated learning to develop a virtual interpreter for ASL, aimed at enhancing accessibility and communication for the deaf community in smart cities. The proposed paradigm establishes a baseline to facilitate the creation of innovative applications that generate contextually and grammatically accurate sign language interpretations. By employing federated deep learning, the framework maintains user privacy while allowing for continuous improvement of the interpreter's performance. This paradigm aims to promote inclusivity and assistive technologies by integrating sign language into urban technological solutions.

Keywords- Arabic sign language; Federated deep learning; Accessibility; Avatar; Smart city.

1. Introduction

Since Smart cities rely on Internet of Things (IoT) sensors for data collection and support a range of applications across fields such as public services, resource management, and communication (Zheng et al., 2022). Additionally, smart cities offer effective solutions for key challenges such as IoT development (Li et al., 2020), healthcare (Ghazal et al., 2021), transportation (Ushakov et al., 2022), and communication systems (Guan et al., 2018). Throughout this extensive data exchange process, sensors produce vast amounts of data. Deaf and Hard of Hearing (DHH) individuals who primarily use sign language often consider themselves part of a linguistic and cultural minority, as they lack full access to the same linguistic resources available to hearing individuals (Bramwell et al., 2000).

Traditional sign language interpretation services and accessibility measures are usually not able to meet the needs of DHH individuals, especially in non-Western regions (Othman et al., 2024). In smart cities, AI-powered signing avatars present a promising solution to bridge this accessibility

gap. These avatars can enhance the inclusion and visibility of sign language within digital content and public services, creating more accessible urban environments. By integrating signing avatars in smart city infrastructure, DHH individuals can experience improved communication access in areas like transportation, healthcare, and public information, reinforcing the cultural and linguistic identity of Deaf communities globally.

Federated learning is a decentralized approach to machine learning that enables multiple devices or entities to collaboratively train a shared model while keeping their data locally stored (Kairouz et al., 2021). Using avatars in smart cities that leverage federated learning enables them to become more adaptive to user needs while maintaining user privacy. This way each avatar functions locally on a user's device or within a specific setting, such as a public kiosk or a healthcare center, and works exclusively with its own unique data. For instance, avatars serving DHH users can process local information to respond to user queries, interpret sign language, or offer guidance in navigation, all while storing this sensitive interaction data locally.

In this paper, we introduce a preliminary model for a realistic Arabian avatar that interprets Arabic Sign Language (ArSL) in smart city ecosystems, preserving efficiency, accuracy, privacy, and security. The FL-based avatar eliminates the need to share raw data with a central server. Instead, they process local data to learn and adapt, updating their models to align with each user's preferences and needs. Each avatar periodically sends model updates, such as updated weights or features, to a central aggregator. The aggregator combines these updates to refine a global model, which is then distributed back to the avatars. This enables them to incorporate the collective knowledge of all users while ensuring that personal information is never exchanged. This collaborative technique allows each avatar to improve in accuracy and contextual awareness over time. For example, an avatar serving DHH users in a hospital can learn how to interpret medical instructions, whereas an avatar in a transportation hub can enhance its ability to provide real-time navigation help. As a result of the federated learning process, avatars can gain knowledge in a variety of circumstances, benefiting from experiences in multiple environments.

The remaining part of this paper is organized as follows: Section 2 highlights the key benefits of employing FL for ArSL, Section 3 reviews the related avatar-based works, Section 4 presents the structure of proposed ArSL avatar within smart cities, Section 5 illustrates the avatar's communication framework, Section 6 discusses the experimental simulation results, and Section 7 concludes this work.

2. Benefits of Federated Learning for ArSL in Smart City

Adopting the federated learning paradigm for Arabic sign language in smart cities provides numerous advantages, including enhanced performance, accessibility, real-time adaptation, strengthened privacy, scalability, and collaborative learning. Figure 1 illustrates the advantages of using FL-based avatars in smart cities, which are further discussed in the following subsections.

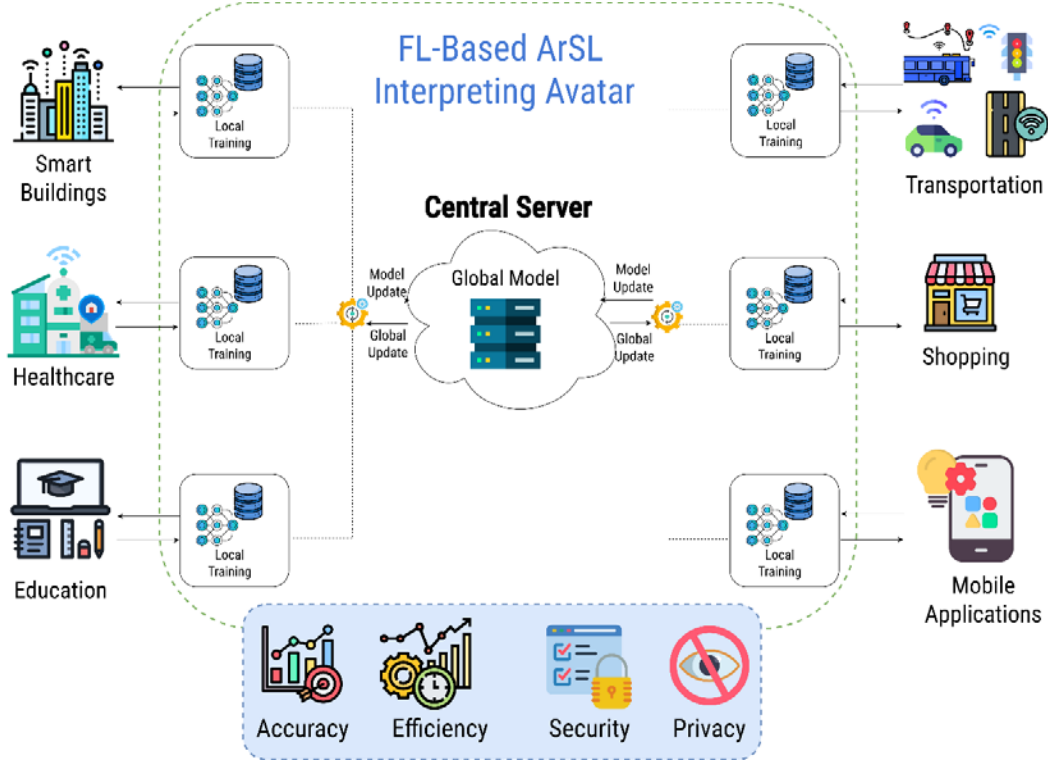


Figure 1. A generic federated learning paradigm within smart cities.

2.1. Enhanced Interpretation Accuracy

This refers to the avatar's ability to provide responses that are relevant, precise, and contextually appropriate for each user's needs (Garcia et al., 2023). An accurate avatar can understand user intent, whether interpreting sign language, offering emergency assistance, or helping with city navigation. Contextual adaptation allows the avatar to customize its responses based on the user's location, recent interactions, or unique needs. Federated learning improves accuracy by enabling the avatar to learn from local interactions across various devices rather than centralizing data (Pandya et al., 2023). However, several issues can significantly impact the effectiveness of various applications in smart cities, such as traffic management and public safety. Factors like data heterogeneity, sparse data availability, variable client engagement, and resource constraints of IoT devices contribute to these challenges. Inconsistent data from diverse sources, combined with dynamic environments that can change rapidly, often leads to non-IID data distributions and gaps in training data, resulting in models that do not accurately reflect city conditions.

The proposed FL-based avatar aims at addressing these challenges by enhancing data integration through standardizing inputs from various sources and implementing an adaptive decentralized learning technique that adjusts models based on local conditions. Additionally, it facilitates hierarchical clustering of clients based on geographical regions or services, allowing for more

targeted model training. Furthermore, it utilizes robust aggregation techniques to enhance the resilience of the global model, ensuring it remains accurate despite the diverse and dynamic nature of data sources within smart cities. Additionally, the avatar's ability to recognize and accurately interpret Arabic dialect enhances its overall accuracy by improving its responsiveness to the unique needs of the Arabic-speaking deaf community. This continuous improvement process enhances the model's understanding, resulting in more accurate responses. Real-time updates enable the avatar to react to changing situations, such as providing timely weather or traffic updates based on actual city conditions, thereby boosting guide accuracy.

2.2. Improved Real-Time Adaptation and Accessibility

The goal is to deliver fast, resource-efficient, and responsive assistance to users. This is accomplished by integrating edge computing with federated learning, enabling avatars to function with low latency, reduced bandwidth consumption, and optimized resource utilization. By processing data on local edge devices, such as nearby servers or user devices, the FL-based avatar avoids delays from central cloud processing and enables faster responses and real-time feedback. Data on local conditions, such as traffic or public transit schedules, is instantly accessible, allowing the avatar to quickly provide the fastest route home if needed. However, communication overhead may arise in such scenarios due to the potentially large size of model updates sent to the central server, particularly when dealing with complex models. Additionally, a large number of clients can create overhead, as each client sending its updates may significantly increase network traffic. Data heterogeneity may also contribute to this issue, as variations in data across clients can lead to differences in update sizes and frequencies, complicating the aggregation process.

The proposed avatar addresses these challenges and achieves high efficiency by minimizing the size of model updates (e.g., weights and gradients) sent to the central server, thereby substantially reducing bandwidth requirements. Additionally, the use of asynchronous FL (Xu et al., 2023) allows updates from clients to be processed individually while reducing waiting time and preventing network traffic. This approach to selective data sharing ensures the avatar remains lightweight and efficient while leveraging shared user experiences across the city.

Federated learning thus makes it possible for ArSL models to be continuously improved through local adaptation and distributed user feedback. For those who use sign language, this creates more inclusive smart city settings that improve interaction and communication. Moreover, the decentralized nature of federated learning allows the model to adapt to regional dialects, user preferences, and specific needs in real time, making it more effective in diverse urban settings.

2.3. Maintain Privacy and Security

Federated learning's capacity to maintain privacy is one of its main benefits. This is achieved by using FL and edge computing among dispersed avatars in smart cities. This setup allows avatars to process and learn from local data within their designated areas, ensuring that sensitive

information remains within the local environment rather than being transmitted to a central server. Information leakage, however, is a major concern in smart cities that depend on FL systems since hackers may use strategies like model poisoning or data poisoning to take advantage of vulnerabilities. Additionally, algorithms such as gradient descent can leak sensitive user information, further exposing the FL ecosystem to potential attacks (Pandya et al., 2023).

The proposed FL-based communication paradigm enables distributed avatars with privacy-preserving techniques that only share anonymized model updates rather than raw data, thus reducing data exposure. Moreover, encrypted communication channels between avatars protect all interactions, while decentralized data management allows each avatar to control data locally, enhancing privacy. This technique not only respects privacy rules, but it also preserves user trust by assuring that sensitive data is handled securely on the local level. Furthermore, the proposed FL framework maintains security by restricting the impact of potential attacks to specific areas rather than affecting the entire system, where regular updates and security patches can be given directly to each avatar, assuring their continued protection without relying on a central system for critical security functions.

3. Related Work

Various avatar systems have been developed to enhance accessibility in education, communication, and real-time interactions for individuals with disabilities, particularly within the Arabic-speaking community. These avatars vary in functionality, focus, and underlying technologies. Only a few of these avatars are explicitly designed to address communication issues with people who are deaf or hard of hearing.

Mind Rockets (Mindrockets, 2024) offers avatars for translating text into Arabic sign language and other sign languages. These avatars are lightweight and easily integrated into websites, focusing on accessibility for the deaf and other individuals with disabilities, such as those with Attention-Deficit/Hyperactivity Disorder (ADHD) or dyslexia. Mind Rockets avatars' primary advantages are their broad covering of disabilities, ease of integration, and compatibility with websites, mobile apps, and other digital platforms. It can only translate text to signs, though.

Eshara avatar system (Eshara, 2024) is another significant advancement in ArSL translation. This system translates Arabic sign language into written or spoken text and vice versa scaling across multiple dialects. Eshara focuses on bidirectional communication, making it not only capable of translating sign language gestures into text but also interpreting text into sign language via an animated avatar. Only a web application is now available for this avatar system, which is still relatively new.

BuHamad avatar (Othman & El Ghouli, 2022) is a Qatari innovation that translates text into Qatari sign language. With its culturally significant attire (Ghutra and Thobe) and highly realistic

animations, the avatar resonates deeply with the local deaf community. It is supported by a sophisticated cloud-based architecture that enables efficient real-time rendering even in low-bandwidth environments. It prioritizes cultural relevance and accessibility, serving as a bridge between the hearing and deaf communities within Qatar. Its architecture includes components for translating Arabic text into sign language animations and a database of annotated signs and sentences. Its realistic design, cloud efficiency, and cultural integration make it remarkable. The fact that this avatar is only available in the Qatari dialect, with possible scalability issues for other Arabic-speaking regions, is one of its primary limitations.

Our proposed FL-based avatar provides notable advancements over existing systems, including enhanced privacy, greater dialect flexibility, improved accessibility, and better adaptability across various sectors. Unlike traditional centralized systems that process user data on cloud platforms and expose sensitive information, our avatar is trained on decentralized data through federated learning. This ensures robust privacy preservation while constantly improving translation capabilities. It is also designed to interact seamlessly across various platforms and sectors within smart cities, setting a new standard for inclusive communication and accessibility for the Arabic-speaking deaf community. In comparison, Eshara, BuHamad, and Mind Rockets offer strong localized solutions but may lack the broad adaptability and privacy aspects provided by our avatar. It is also designed for adaptive and context-aware interactions, as it customizes its appearance based on the specific sector (e.g., medical, customer service, transportation). This ensures that our avatar is not only linguistically correct but also context-appropriate. Existing systems employ static designs that lack this sector-specific adaptability, reducing their effectiveness in diverse smart city applications.

Table 1. A summary of FL-based avatar compared to the related ArSL avatars.

	BuHamad	Mind Rockets	Eshara	FL-Based Avatar
Target Audience	Qatari Deaf Community	General & Arabic Deaf Community	Arabic Deaf Community	Arabic Deaf Community
Dialect Support	Limited to Qatari Sign Language	Multiple dialects and languages	Limited dialect support	Diverse Arabic dialects and idioms
Cultural & Regional Relevance	Limited to Qatari Sign Language	Multiple dialects and languages	Limited dialect support	Diverse Arabic dialects and idioms
Privacy & Security	Strong cultural relevance	Limited cultural customization	Limited cultural relevance	Adaptive design, region-specific uniforms
Facial Expressions & Gestures	Fixed avatar design for Qatari context	Fixed avatars adaptable to various platforms	Customizable avatars	Adaptive avatars based on sector
Platform Compatibility	Realistic 3D gestures	Standard sign language gestures	Standard sign language gestures	Expressive facial features and nuanced gestures

4. The Avatar Structure

The proposed FL-based framework involves a realistic Arabian avatar created to facilitate ArSL translation in smart city ecosystems with a focus on efficiency, accuracy, and privacy. Translating ArSL provides unique issues due to its complicated linguistic structure and the limited resources accessible for ArSL linguistics, particularly when dealing with varied Arabic dialects and regional idioms. Existing avatar-based systems frequently rely on a crude word-by-word translation technique, which fails to convey the nuances of actual language expression required for ArSL and lacks proper dialect support (Othman & El Ghouli, 2022). To address these challenges, the avatar in this work utilizes a federated learning approach, enabling it to enhance its translation abilities by training on diverse data sources while preserving privacy. This aims at improving the ability to recognize and accurately interpret various Arabic dialects, hence increasing responsiveness to the specific demands of the Arabic-speaking deaf community.

The potential of FL-based avatar extends to a variety of smart city sectors, including public transportation, healthcare, customer service, education, and more. Where it can facilitate more inclusive and accessible interactions for deaf people. In these sectors, the avatar's appearance will adapt to suit the context, enhancing reliability and realism in communication. For instance, in healthcare settings, the federated learning-based avatar will wear a medical uniform; in customer service environments, it will be dressed in formal attire; and for transportation, a casual uniform will be used. This adaptive design aims to offer a more immersive and respectful experience for users across various applications.

Several technical problems must be overcome to ensure that our proposed avatar can meet the needs of real-time and realistic sign language translation. The avatar must demonstrate lifelike animation, including expressive facial features and natural gestures, which are vital for conveying subtleties and emotions, as shown in Figure 2. Furthermore, the FL-based avatar is intended to work seamlessly across a variety of platforms, including mobile phones, web browsers, and smartwatches, while preserving performance on low-speed internet connections and limited graphical capabilities. Cross-platform compatibility and responsive real-time capabilities are required for successful deployment within the smart city framework.



Figure 2. A demonstration of Avatar's expressive body features.

The proposed FL-based avatar provides a promising improvement in avatar-based sign language translation by combining advanced linguistic modeling and federated learning, addressing the objectives of privacy, dialectic correctness, and efficient operation that smart city applications require. This FL-based ArSL paradigm aims to set a novel standard for accessible communication by fostering a more inclusive urban environment for the Arabic-speaking deaf community. To animate the 3D avatar, we use two different components that improve its visual depiction, as follows:

Skeleton-Based Animation. We present a skeleton structure for the FL-based ArSL avatar, which allows for exact control of bodily movements. We meticulously skinned the avatar's mesh to match the stiff transformations of each bone in the skeleton. This technique ensures that when the rig is adjusted, the matching mesh deforms smoothly, allowing for natural body and hand gestures that are essential for efficient communication in ArSL systems.

Vertex Morphing Animation. For detailed facial animations, we utilize vertex morphing. Figure 3(a) illustrates how to produce expressive face expressions by altering the placements of specific vertices within the avatar's geometry. Specific mathematical functions can be used to save or modify the new position of each vertex. Because of its high resource requirements, this technology is only used for creating sophisticated facial movements. Figure 3(b) displays the 3D object representing the construction of the avatar's face.

The proposed baseline requires a set of sophisticated tools and services to help with the avatar's animation. This design will enable developers to quickly integrate the FL-based avatar into various applications while offering the capability for converting written text into interpreted Arabic sign language.



Figure 3. Avatar's front expression (left) and 3D face (right).

5. The Federated Communication Framework

In this work, the FL-based ArSL Avatar works within a decentralized communication framework architecture, with each sector using its avatar. Each avatar is provided with a local dataset and a local model, allowing it to effectively interact with the deaf community while continuously improving its performance via a federated learning strategy.

5.1. Federated Learning Process

Figure 4 shows the federated learning process. Each sector's avatar i has access to a local dataset D_i containing interaction data with local deaf users. The avatar uses this dataset to train its local model, M_i . The local training process can be structured as follows:

$$M_i^{(t+1)} = M_i^{(t)} - \eta \nabla L(M_i^{(t)}, D_i) \quad (1)$$

where:

- $M_i^{(t+1)} = M_i^{(t)} - \eta \nabla L(M_i^{(t)}, D_i)$
- η is the learning rate.
- L represents the loss function computed on the local dataset D_i .

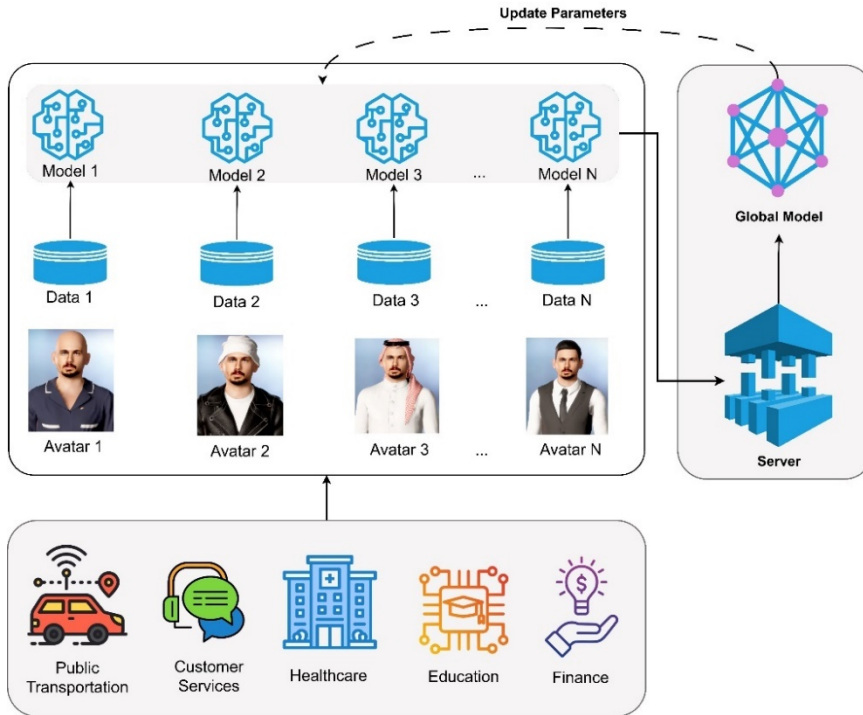


Figure 4. Avatar communication in federated deep learning.

After completing the local training, the avatar generates a model update ∇M_i , which captures the changes made to its parameters:

$$\nabla M_i = M_i^{(t+1)} - M_i^{(t)} \quad (2)$$

At the end of a predefined period (e.g., daily), all local model updates from the avatars in various sectors are sent to a central server for aggregation. The aggregation process can be represented as:

$$M^{(t+1)} = M^{(t)} + \frac{1}{N} \sum_{i=1}^N \nabla M_i \quad (3)$$

where:

- $M^{(t)}$ are the global model parameters at iteration t .
- N is the total number of avatars (local models) participating in the update.

The aggregated update is then used to refine the global model M at the server:

$$M^{(t+1)} = M^{(t)} + \alpha(M^{(t+1)} - M^{(t)}) \quad (4)$$

where α is a hyperparameter that determines the contribution of the aggregated update to the global model. After the global model has been updated, the new parameters are distributed back to each avatar, allowing them to initialize their local models with the improved global knowledge. This process is done iteratively, with each avatar continuously learning from its interactions with local deaf users, updating its local model, and contributing to the global aggregate. This results in a steadily evolving avatar system capable of better understanding and interpreting Arabic sign language. The federated learning approach allows the FL-based Avatar in smart cities to maintain users' privacy while enhancing communication capabilities. By exploiting local data for personalized learning and collecting model updates regularly, avatars may easily adapt to the demands of the deaf community in a variety of industries, resulting in improved communication and inclusion in smart urban environments.

5.2. Federated Learning Infrastructure in Real-World Scenarios

Implementing the federated learning paradigm in real-world applications requires a robust, scalable infrastructure that is tailored to the specific needs of each sector while adhering to common principles. FL is based on edge computing, with intelligent edge nodes equipped for local training and inference, as shown in Figure 5. These nodes use modern hardware, such as CPUs, GPUs, and specialized processors such as TPUs, to do complex machine-learning tasks on-site, lowering latency and increasing responsiveness (Duan et al., 2023). For example, IoT-enabled medical equipment and localized servers in healthcare allow for sensitive patient data training, whilst onboard computer units in cars provide real-time processing of traffic and sensor data in

public transit. Similarly, devices such as tablets and laptops when combined with institutional servers provide personalized learning in education, while branch servers, ATMs, and consumer devices enable localized FL models in customer service and finance.

High-capacity storage systems are critical for managing massive, dispersed datasets across different industries, but data privacy is protected by keeping it localized at edge nodes (Salh et al., 2023). Secure on-premises storage solutions are critical, including encrypted Solid-State Drives (SSDs) for medical records, localized storage for traffic and passenger data, distributed academic datasets from learning management systems, customer interaction logs for service models and encrypted financial data for fraud detection. Furthermore, efficient and secure communication networks, such as 5G, fiber-optic, or Wi-Fi, are required for continuous model updates between edge nodes and central aggregation hubs (Yang et al., 2023). These networks allow nodes to safely communicate model parameters while maintaining data privacy.

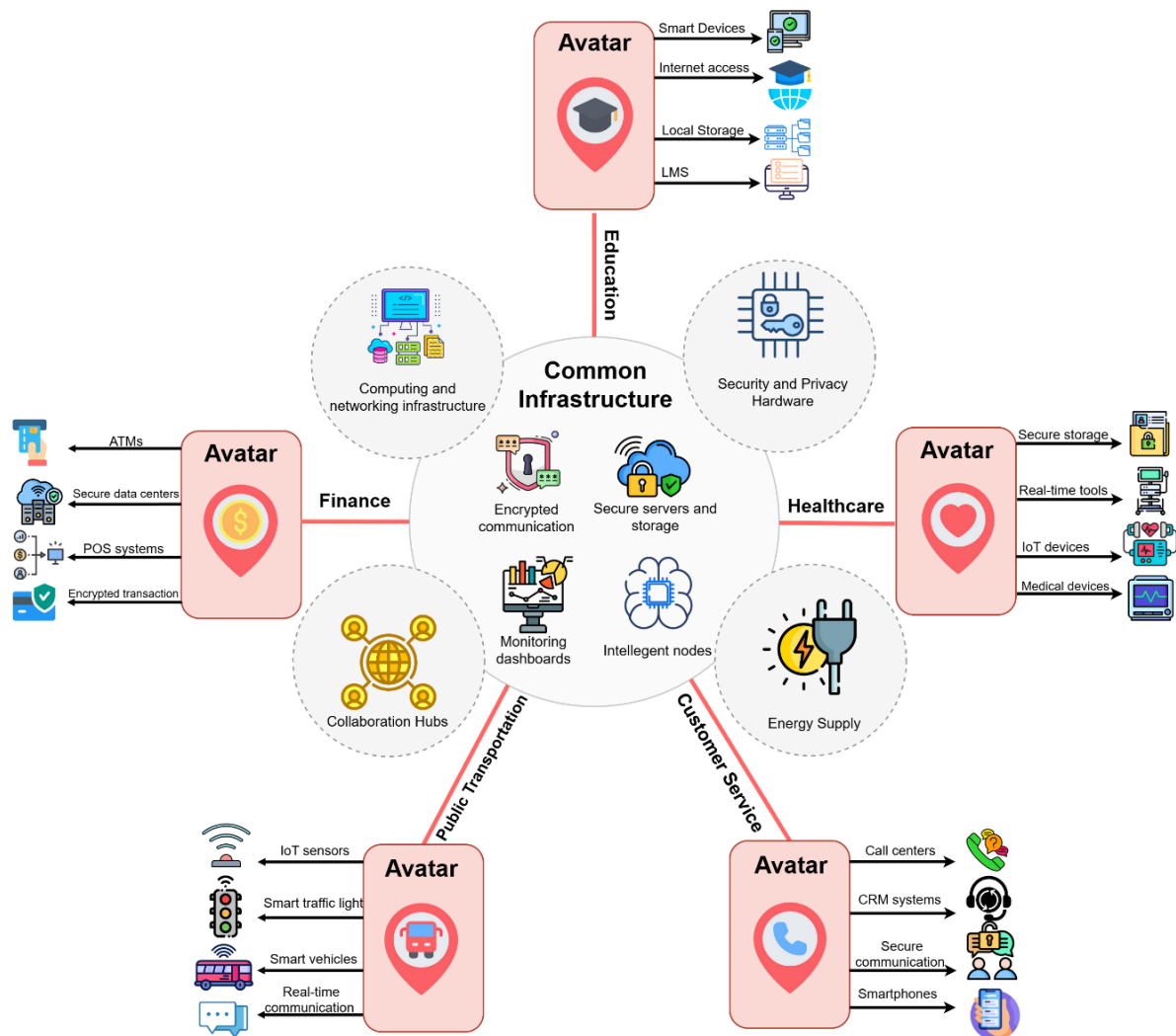


Figure 5. Real-time infrastructure of avatar-based ArSL interpreter in smart cities.

The execution layer of FL infrastructure supports a broad range of computing devices based on unique use cases, including high-performance GPUs for training sophisticated diagnostic models, ruggedized CPUs for dynamic situations, and lightweight edge devices for smaller models, as shown in Figure 5. These devices require dependable power supply, with backup systems such as UPS or generators assuring continuous operation. Efficient systems keep servers running effectively, while ruggedized hardware enables operations in tough situations such as distant educational settings or transportation systems.

Centralized coordination is critical to achieving interoperability, scalability, and compliance within FL systems. A central hub implements established protocols for model integration, providing seamless collaboration among edge nodes and sectors. Regulatory compliance is crucial, including frameworks such as GDPR for financial services and HL7 standards for healthcare. Dashboards provide real-time monitoring of system performance, node involvement, and scalability, enabling more effective administration.

The adaptability of FL infrastructure is proven in a variety of domains. In healthcare, FL allows for collaborative training of diagnostic models across institutions while protecting sensitive patient information. FL enhances public transportation by allowing for real-time optimization of traffic patterns and route scheduling based on vehicle data. The education industry uses FL to construct personalized learning models based on dispersed datasets from numerous universities. FL improves customer experience by creating adaptable models that include feedback from remote contact centers. In finance, FL helps to identify fraud by allowing institutions to collaborate on transaction data, boosting accuracy while retaining secrecy. The FL-based avatar illustrates the power of decentralized, privacy-preserving machine learning by addressing sector-specific concerns, and fostering industry-wide innovation while assuring compliance, security, and efficiency.

6. Experimental Results

6.1. Simulation Setup

In this paper, VGG19 and VGG16 (Simonyan, 2014) deep learning models are used as backbone for the federated deep learning framework. Several experiments are conducted to select the best hyperparameters for this model. The performance is evaluated on ArASL2018 dataset (Latif et al., 2019) in terms of average training time, accuracy, precision, recall, and f1-score. Each experiment uses 64 data samples, with five clients (representing 5 avatars) trained over ten epochs. The federated averaging method aggregates gradients on the server side, while categorical cross entropy is used as a loss function. Softmax is used as an activation function, and stochastic gradient descent is used as a primary optimizer with a learning rate of 0.01.

6.2. Performance on ArSL Recognition

Table 2 summarizes the performance results of all FL-based models. This evaluation procedure provides a detailed understanding of the model's performance across various avatar's classification characteristics, ensuring its ability to handle various scenarios using several data splits. However, due to the imbalanced nature of the Arabic sign language dataset model's performance is assessed using accuracy and macro-averaging metrics. In imbalanced circumstances, macro-averaging is very useful since it treats all classes equally without being affected by the majority class. The FL-VGG19 has a high testing performance, achieving an accuracy of 98.8%, a precision of 98.79%, a recall of 98.78%, and an F1-score of 98.78%.

Table 2. Performance of VGG16 and VGG19 on avatar simulation dataset.

Model	Val. Acc	Test Acc.	Precision	Recall	F1-Score	Training time per client (sec.)	Training time per round (min.)
VGG16	97.30	98.70	98.72	98.71	98.71	60	5
VGG19	97.10	98.80	98.79	98.78	98.78	65	5.4

However, FL-VGG16 and FL-VGG19 show comparable performance across all metrics. FL-VGG16 achieves slightly higher validation accuracy, with a marginal difference of 0.2%, underscoring the minor discrepancies between the performances of the two models. VGG16 takes 60 seconds per avatar for local data training, while VGG19 takes 65 seconds. The additional layers in VGG19 slightly increase its computational requirements. On the other hand, VGG16 requires 5 minutes per training round, while VGG19 takes 5.4 minutes, reflecting the added computational complexity of VGG19. Overall, VGG19 offers slightly better performance metrics but at the cost of increased training time. The differences, however, are minimal, suggesting that either model could be suitable depending on the specific requirements of the application.

7. Conclusion

This paper presents a foundational framework for creating an AI-powered virtual interpreter that can generate and interpret ArSL using a federated learning approach. This innovative FL-based avatar demonstrates realistic hand movements and authentic facial expressions and prioritizes privacy and data security by leveraging federated deep learning paradigm. The virtual interpreter can be effectively deployed in a smart city environment, enabling smart city technologies to be more inclusive and accessible. This proposed framework lays a groundwork for further enhancing the integration of sign language interpretation into various applications, ultimately fostering improved communication and interaction for the deaf and hard-of-hearing population in urban settings.

References

- Bramwell, R., Harrington, F., & Harris, J. (2000). Deafness—disability or linguistic minority? *British Journal of Midwifery*, 8(4), 222–224.
- Duan, Q., Huang, J., Hu, S., Deng, R., Lu, Z., & Yu, S. (2023). Combining federated learning and edge computing toward ubiquitous intelligence in 6G network: Challenges, recent advances, and future directions. *IEEE Communications Surveys & Tutorials*.
- Eshara: Arabic Sign Language Translation. (2024). Retrieved from <https://bit.ly/41kZ2EP>
- García-Luque, R., Toro-Gálvez, L., Moreno, N., Troya, J., Canal, C., & Pimentel, E. (2023). Integrating Citizens' Avatars in Urban Digital Twins. *Journal of Web Engineering*, 22(6), 913–938.
- Ghazal, T. M., Hasan, M. K., Alshurideh, M. T., Alzoubi, H. M., Ahmad, M., Akbar, S. S., ... Akour, I. A. (2021). IoT for smart cities: Machine learning approaches in smart healthcare—A review. *Future Internet*, 13(8), 218.
- Guan, Z., Si, G., Zhang, X., Wu, L., Guizani, N., Du, X., & Ma, Y. (2018). Privacy-preserving and efficient aggregation based on blockchain for power grid communications in smart communities. *IEEE Communications Magazine*, 56(7), 82–88.
- Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... Others. (2021). Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, 14(1–2), 1–210.
- Latif, G., Mohammad, N., Alghazo, J., AlKhalaf, R., & AlKhalaf, R. (2019). ArASL: Arabic alphabets sign language dataset. *Data in brief*, 23, 103777.
- Li, T., Zhao, M., & Wong, K. K. L. (2020). Machine learning based code dissemination by selection of reliability mobile vehicles in 5G networks. *Computer Communications*, 152, 109–118.
- Mind Rockets Inc. (2024). Mind Rockets: Arabic Sign Language Solutions. Retrieved from <https://main.mindrocketsinc.com/en>
- Othman, A., Dhouib, A., Chalghoumi, H., Elghoul, O., & Al-Mutawaa, A. (2024). The Acceptance of Culturally Adapted Signing Avatars Among Deaf and Hard-of-Hearing Individuals. *IEEE Access*.
- Othman, A., & El Ghoul, O. (2022). BuHamad: The first Qatari virtual interpreter for Qatari Sign Language. *Nafath*, 6(20).
- Pandya, S., Srivastava, G., Jhaveri, R., Babu, M. R., Bhattacharya, S., Maddikunta, P. K. R., ... Gadekallu, T. R. (2023). Federated learning for smart cities: A comprehensive survey. *Sustainable Energy Technologies and Assessments*, 55, 102987.
- Salh, A., Ngah, R., Audah, L., Kim, K. S., Abdullah, Q., Al-Moliki, Y. M., ... & Talib, H. N. (2023). Energy-efficient federated learning with resource allocation for green IoT edge intelligence in B5G. *IEEE Access*, 11, 16353–16367.
- Simonyan, K. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

- Sun, T., Li, D., & Wang, B. (2022). Decentralized federated averaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4), 4289–4301.
- Ushakov, D., Dudukalov, E., Mironenko, E., & Shatila, K. (2022). Big data analytics in smart cities' transportation infrastructure modernization. *Transportation Research Procedia*, 63, 2385–2391.
- Xu, C., Qu, Y., Xiang, Y., & Gao, L. (2023). Asynchronous federated learning on heterogeneous devices: A survey. *Computer Science Review*, 50, 100595.
- Yang, C., Chen, Y., Zhang, Y., Cui, H., Yu, Z., Guo, B., ... & Yang, Z. (2023). RaftFed: a lightweight federated learning framework for vehicular crowd intelligence. *arXiv preprint arXiv:2310.07268*.
- Zheng, Z., Zhou, Y., Sun, Y., Wang, Z., Liu, B., & Li, K. (2022). Applications of federated learning in smart cities: recent advances, taxonomy, and open challenges. *Connection Science*, 34(1), 1–28.