Automatic Gesture-Based Arabic Sign Language Recognition: A Federated Learning Approach

Ahmad Alzu'bi[1], Tawfik Al-Hadhrami[2], Amjad Albashayreh[1], Lojin Bani Younis[1]

[1]*Department of Computer Science, Jordan University of Science and Technology, Irbid, Jordan*
agalzubi@just.edu.jo, amalbashayreh20@cit.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

Abstract- Featuring machine learning algorithms for recognizing hand gesture patterns adjusted for individuals with disabilities is an expanding trend in assisted living. This paper addresses the challenge of interpreting the semantics of image-based hand gestures by introducing a federated deep learning architecture for Arabic sign language recognition. The proposed model manages distributed learning through a client-server paradigm, wherein several edge nodes collaborate to jointly learn the discriminative features of confidential data without breaching its privacy. This model will enable more accessibility for people with deafness or impairment using image gestures. The federated learning procedure is primarily based on the ResNet32 deep backbone and federated averaging mechanism. The experimental results show the effectiveness of the proposed FL model, achieving an accuracy of 98.30% with 33 seconds on average for each client in a single training round. This demonstrates its high capabilities in recognizing Arabic sign language and improving the communication experience for people with disabilities.

Keywords- Arabic sign language; Federated deep learning; Image recognition; Accessibility; Communication disabilities.

1. Introduction

Since sign language serves as the main method of communication for millions globally, there's considerable enthusiasm surrounding the potential uses of advanced Sign Language Recognition (SLR) tools (Semreen, 2023) (Al-Qurishi et al., 2021). Given the diverse array of opportunities, these assistive technologies could extend beyond mere translation. They could enable accessible sign language broadcasts, promote the creation of responsive devices capable of seamlessly interpreting sign language commands, and even spearhead the development of intricate systems tailored to aid individuals with impairments in accomplishing daily tasks with greater autonomy (Othman et al., 2024).

People with disabilities, such as those who are deaf or hard of hearing, utilize Sign Language (SL), a visual communication method that uses gestures, facial expressions, and body movements. Leveraging deep neural network architectures, deep learning algorithms analyze vast amounts of data to learn intricate patterns and features inherent in hand movements

(Rastgoo et al., 2021) (Cui et al., 2019). However, there are several issues with image-based SLR systems, particularly concerning the intricacies of feature learning and image processing, the confidentiality of private information, and the effectiveness of SLR systems in practical settings. As a result, it is still very important to maintain the speed, accuracy, and reliability of interpretation algorithms (Elsheikh, 2023) (Cheok et al., 2019).

Federated Learning (FL) is an emerging machine learning paradigm associated with decentralized methods, proving to be an effective approach for training shared global models (Wen et al., 2023). FL methods entail coordinating the training of a central model from a collection of participating devices. When training data is sourced from user interactions with mobile applications, for instance, one significant application scenario for FL arises (Lee et al., 2024). In this context, FL enables mobile phones to collectively learn a shared prediction model while retaining all training data on the device, effectively performing computations on their local data to update a global model. This approach goes beyond the use of local models for mobile device predictions by bringing model training to the device level. Within the context of SLR, this approach provides a promising solution to the challenges of privacy preservation, data diversity, and model adaptability (Krishnan and Manickam, 2024) (You et al., 2023).

Arabic sign language (ArSL) encompasses a rich vocabulary and intricate structures. Much like other languages, it involves the combination of hand shapes, orientations, motion, and facial expressions to convey various meanings (Zakariah et al., 2022). While various deep learning algorithms have been applied to recognize Arabic sign language (Aldhahri et al., 2023) (Saleh and Issa et al., 2020) (Ahmed et al., 2021) (Kamruzzaman et al., 2020) (Alawwad et al., 2021), prior studies did not employ federated learning architectures. This motivated us to address this gap by utilizing and investigating a federated deep learning model to recognize the Arabic sign language, ensuring privacy for individuals with disabilities and providing high performance with low time complexity. This allows model training to take place locally on the local devices of users or decentralized servers, protecting the privacy of confidential information. Therefore, this approach enables more accurate and robust recognition of gestures across diverse environments and conditions.

The rest of this article is organized as follows: Section 2 presents the procedure of image preprocessing; the proposed architecture of the FL-based model is introduced in Section 3; Section 4 presents the experimental results; Section 4 discusses model applicability, scalability, and ethical issues; and Section 5 concludes this study.

2. ARASL Images Preparation

The benchmarking dataset utilized in this study is the Arabic Alphabet Sign Language (ARASL) dataset (Latif et al., 2019), which consists of 54,049 images depicting hand gestures representing the Arabic alphabet. This dataset is specifically designed to assist the deaf community in

understanding the language and expressing their thoughts and emotions freely. Comprising 32 classes corresponding to Arabic letters, each class contains a specific number of images. Figure 1 displays a selection of sample ARASL hand-gesture images.

A transformation procedure was applied to ARASL data consisting of image resizing, tensor conversion, and [0,1] normalization. Finally, the image collection is divided into 70% for training, 10% for validation, and 20% for testing. Multiple subsets of the training and testing images are created, which is necessary for simulating different decentralized clients in a FL framework.

Figure 1. Sample images of Arabic signs from ARASL dataset.

3. Method

Figure 2 illustrates the general framework of the proposed federated learning architecture, which includes a central server interacting with multiple clients functioning as distributed computing nodes. The server hosts a global deep learning model designed to be trained on the local data of the clients. On the client side, each client holds a subset of Arabic hand-gesture images containing labeled samples. To maintain privacy, clients do not share their local ASL images with the server or other clients. The server initially broadcasts the global model to all participating clients, utilizing data from each client collaboratively. This process aims to identify the optimal model weights that minimize the classification loss rate for each client. Over several training rounds, the server aggregates the training results, which represent the gradients of the local model parameters, updates the global model, and then sends it back to the clients.

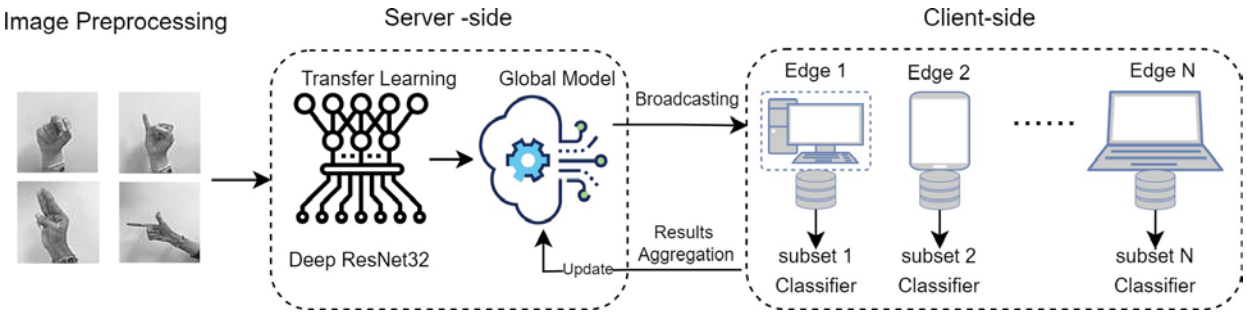

Figure 2. The pipeline of the federated learning process for Arabic sign recognition.

In this study, the Federated Averaging (FedAvg) (McMahan et al., 2017) is used for data aggregation with a network of five clients, utilizing Distributed Stochastic Gradient Descent (D-SGD). The training process involves 10 local epochs and 10 global rounds with iterative model updates. This approach synchronizes the local contributions of each client, leading to enhanced global hand-gesture image classification. The server continuously updates the global model after each round and redistributes these updates to the local models on the client side.

ResNet32 (He et al., 2016), a well-recognized deep neural network architecture, has been incorporated into our federated framework to facilitate the training and evaluation processes across a network of participating client devices. This approach enables efficient transfer learning from a general domain to the specific ArSL domain.

4. Experimental Results

4.1. Experiments Setup

To determine the optimal hyperparameters for evaluating the FL model's performance, several experiments are carried out. In every experiment, five clients perform ten epochs of training on local data. Gradients are aggregated using FedAvg on the server side, and the architecture is configured with categorical cross-entropy loss function, SoftMax function for image classification, SGD optimizer, and a learning rate of 0.01. The classification accuracy of ArSL image recognition is calculated. The true and false measurements (TP, TN, FP, and FN) are used to compute standard evaluation metrics such as accuracy, precision, recall, and F1-score. High accuracy reflects the model's effectiveness in correctly identifying various hand movements and reducing classification errors.

4.2. ASL Recognition Results

Figure 3 presents the macro-average results of the proposed FL-ResNet32 model over ten rounds. The FL-ResNet32 demonstrates consistently high performance in both testing and validation, achieving a test accuracy of 98.3%, precision of 98.28%, recall of 98.26%, and an F1-score of 98.27%. Accuracy and macro-average metrics are employed to assess the model's

performance, particularly because the Arabic sign language dataset is imbalanced. Macro-averaging treats all classes equally without favoring the dominant class.

In terms of training time, FL-ResNet32 effectively recognizes Arabic sign language with an accuracy of 98.3% in an average of 33 seconds over 10 epochs. Additionally, the entire model training across 10 rounds with 5 distributed clients (edge nodes) takes approximately 28 minutes on average.

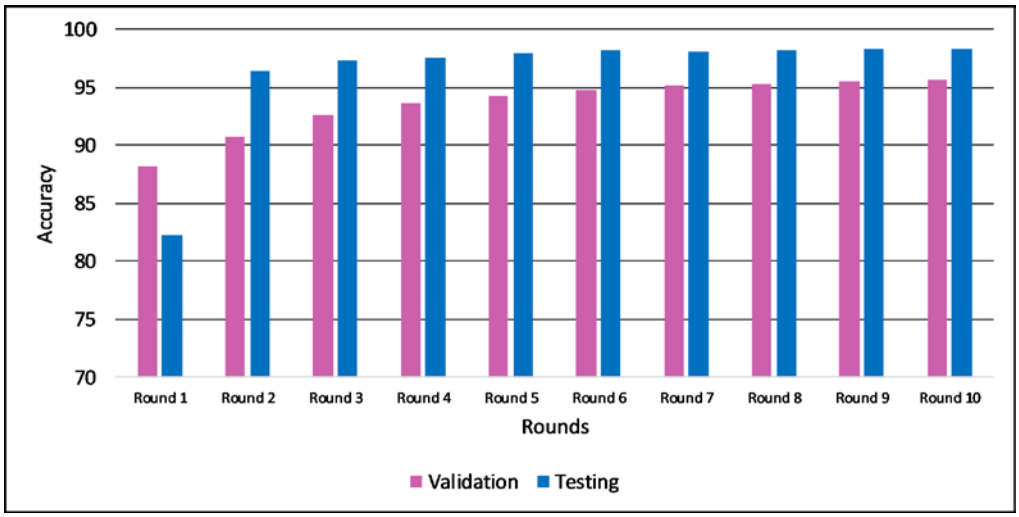

Figure 3. Macro Average accuracy achieved by the FL-ResNet32 on ArSL images.

Table 1 provides a performance comparison between our proposed federated deep learning model and existing ArSL recognition approaches evaluated on the ArASL2018 dataset. The table highlights key features and performance results documented during testing. As shown, FL-ResNet32 outperforms the other methods and recognizes ASL images more accurately, achieving an accuracy of 98.3% in an average of 33 seconds over 10 epochs.

Table 1. Comparison of macro average accuracy on ArASL2018 with related works.

Research Ref.	Method	Test Accuracy (%)	Epochs
Kamruzzaman et al. (2020)	CNN	90.0	100
Aldhahri et al. (2023)	MobileNet	94.5	15
Zakariah et al. (2022)	EfficientNet-B4	95.0	30
This Work	FL-ResNet32	98.3	10

5. Discussion

This study emphasizes how crucial it is to have federated computing environments to enable the utilization of diverse information that can be gathered from various kinds of computing edges, or client devices. This information is extremely sensitive and confidential since it pertains to individuals with disabilities. Conventional machine-learning techniques frequently entail

compiling data on a single workstation or server. But because human communication is so sensitive, privacy concerns must be addressed, especially in Internet of Things (IoT) setups.

However, transferring these data requires a network connection with sufficient bandwidth for large datasets and low latency to ensure timely predictions (Diaz et al., 2023). Additionally, network communication dependency requires sophisticated encryption techniques to ensure privacy and security of sensitive information. Techniques like data compression can be also employed to enhance communication efficiency and increase scalability of FL-based ASL recognition systems.

To facilitate interaction between the deaf community and society, creating a sign language interpreter able to convert sign language into text or spoken language is crucial. This interpreter can be created through computer vision focused approaches enabled in mobile devices (Talov, 2022). To develop a practical and effective system for sign language interpretation, further research in this area is still needed. Recent vision-centric research and systems (Othman et al., 2024) (Othman and El Ghoul, 2022) (Bennbaia, 2022) shifted toward developing culturally adapted signing avatar technologies. This enables individuals with deaf and hard of hearing to engage with community life, leading to the emergence of more dynamic and adaptable communication approaches.

Virtual human avatars, also known as signing avatars or sign language avatars, are a type of conversational technology that uses a 3-D representation of a person to produce text in any sign language or international sign. The use of sign language avatars is one cutting-edge interactive solution to the problem of sign language content access. This avatar-based technology will leverage federated learning, as the communication model in FL-based systems aligns well with a server-client environment, involving various interactive client devices that can provide the server with additional training data in multiple formats, such as text and audio. Further research is necessary to investigate the feasibility of avatar-based intelligent solutions for sign language recognition and translation within large-scale decentralized networks. This advanced technology could greatly enhance communication in future smart cities.

6. Conclusion

This study presents a federated deep learning approach to recognize and classify Arabic sign language using hand-gesture images. The proposed architecture stands out as a successful strategy for attaining high accuracy while keeping the critical practice of protecting patient data privacy, something that existing ArSLR approaches lack. This collaborative distributed learning approach allows for efficient model training on remote devices. The future investigation seeks to enhance the user experience of Arabic sign language recognition through an interactive user interface on mobile phones. This could facilitate contextual learning of sign expressions for individuals with communication disabilities.

References

Ahmed, M., Zaidan, B., Zaidan, A., Salih, M. M., Al-Qaysi, Z., and Alamoodi, A. (2021). Based on wearable sensory device in 3d-printed humanoid: A new real-time sign language recognition system. Measurement, 168:108431.

Al-Qurishi, M., Khalid, T., and Souissi, R. (2021). Deep learning for sign language recognition: Current techniques, benchmarks, and open issues. IEEE Access, 9:126917– 126951.

Alawwad, R. A., Bchir, O., and Ismail, M. M. B. (2021). Arabic sign language recognition using faster r-cnn. International Journal of Advanced Computer Science and Applications, 12(3).

Aldhahri, E., Aljuhani, R., Alfaidi, A., Alshehri, B., Alwadei, H., Aljojo, N., Alshutayri, A., and Almazroi, A. (2023). Arabic sign language recognition using convolutional neural network and mobilenet. Arabian Journal for Science and Engineering, 48(2):2147– 2154.

Bennbaia, S. (2022). Toward an evaluation model for signing avatars. Nafath, 6(20).

Cheok, M. J., Omar, Z., and Jaward, M. H. (2019). A review of hand gesture and sign language recognition techniques. International Journal of Machine Learning and Cybernetics, 10:131–153.

Cui, R., Liu, H., and Zhang, C. (2019). A deep neural framework for continuous sign language recognition by iterative training. IEEE Transactions on Multimedia, 21(7):1880– 1891.

Diaz, J. S. P., & Garcia, A. L. (2023). Study of the performance and scalability of federated learning for medical imaging with intermittent clients. Neurocomputing, 518, 142-154.

Elsheikh, A. (2023). Enhancing the Efficacy of Assistive Technologies through Localization: A Comprehensive Analysis with a Focus on the Arab Region. Nafath, 9(24).

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778.

Kamruzzaman, M. et al. (2020). Arabic sign language recognition and generating Arabic speech using convolutional neural network. Wireless Communications and Mobile Computing, 2020.

Krishnan, R., & Manickam, S. (2024). Enhancing Accessibility: Exploring the Impact of AI in Assistive Technologies for Disabled Persons. Nafath, 9(25).

Latif, G., Mohammad, N., Alghazo, J., AlKhalaf, R., and AlKhalaf, R. (2019). Arasl: Arabic alphabets sign language dataset. Data in brief, 23:103777.

Lee, J., Solat, F., Kim, T. Y., & Poor, H. V. (2024). Federated learning-empowered mobile network management for 5G and beyond networks: From access to core. IEEE Communications Surveys & Tutorials.

McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In Artificial intelligence and statistics, pages 1273–1282. PMLR.

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).

Rastgoo, R., Kiani, K., & Escalera, S. (2021). Sign language recognition: A deep survey. Expert Systems with Applications, 164, 113794.

Saleh, Y., & Issa, G. (2020). Arabic sign language recognition through deep neural networks fine-tuning. International Association of Online Engineering, 71-83.

Semreen, S. (2023). Sign languages and Deaf Communities. Nafath, 9(24).

Talov, M. C. (2022). SpeakLiz by Talov: Toward a Sign Language Recognition mobile application. Nafath, 7(20).

Wen, J., Zhang, Z., Lan, Y., Cui, Z., Cai, J., & Zhang, W. (2023). A survey on federated learning: challenges and applications. International Journal of Machine Learning and Cybernetics, 14(2), 513-535.

You, C., Guo, K., Yang, H. H., & Quek, T. Q. (2023). Hierarchical personalized federated learning over massive mobile edge computing networks. IEEE Transactions on Wireless Communications, 22(11), 8141-8157.

Zakariah, M., Alotaibi, Y. A., Koundal, D., Guo, Y., Mamun Elahi, M., et al. (2022). Sign language recognition for arabic alphabets using transfer learning technique. Computational Intelligence and Neuroscience, 2022.

