Radar-Based Imaging for Sign Language Recognition in Medical Communication

Raffaele Mineo[⋆] [⋆⋆] [1][0000-0002-1171-5672], Gaia Caligiore[⋆] [2][0000-0002-7087-1819], Federica Proietto Salanitri[1][0000-0002-6122-4249], Isaak Kavasidis[1][0000-0003-4366-5195], Senya Polikovsky[3][0000-0002-6030-1863], Sabina Fontana[4][0000-0003-3083-1676], Egidio Ragonese[1][0000-0001-6893-7076], Concetto Spampinato[1][0000-0001-6653-2577], and Simone Palazzo[1][0000-0002-2441-0982]

[1] Department of Electrical, Electronic, and Information Engineering (DIEEI), University of Catania, Catania, Italy
[2] University of Modena and Reggio Emilia, Modena, Italy
[3] Max Planck Institute for Intelligent Systems, Tubingen, Germany
[4] Department of Humanities (DISUM), University of Catania, Catania, Italy

Abstract. Ensuring equitable access to medical communication is crucial for deaf and hard-of-hearing individuals, especially in clinical settings where effective patient-doctor interaction is essential. In this work, we present a novel radar-based imaging framework for Sign Language recognition (with a focus on the Italian Sign Language, LIS), specifically designed for medical communication. Our method leverages 60 GHz mm-wave radar to capture motion features while ensuring anonymity by avoiding the use of personally identifiable visual data. Our approach performs sign language classification through a two-stage pipeline: first, a residual autoencoder processes Range Doppler Maps (RDM) and moving-target indications (MTI), compressing them into compact latent representations; then, a Transformer-based classifier learns temporal dependencies to recognize signs across varying durations. By relying on radar-derived motion imaging, our method not only preserves privacy but also establishes radar as a viable tool for analyzing human motion in medical applications beyond sign language, including neurological disorders and other movement-related conditions. We carried out experiments on a new large-scale dataset containing 126 LIS signs — 100 medical terms and 26 alphabet letters. Our method achieves 93.6% accuracy, 87.9% sensitivity, 99.3% specificity, and an 87.7% F1 score, surpassing existing approaches, including an RGB-based baseline. These results underscore the potential of radar imaging for real-time human motion monitoring, paving the way for scalable, privacy-compliant solutions in both sign language recognition and broader clinical applications. The code is available at https://github.com/IngRaffaeleMineo/SignRadarClassification_MICCAI2025 and the dataset will be released publicly.

Keywords: Deep learning · Sign language · Medical communication.

[⋆] These authors contributed equally to this work.

[⋆⋆] Corresponding author: R. Mineo (raffaele.mineo@unict.it)

1 Introduction

Sign languages (SLs) are complete visual-gestural systems used by deaf communities worldwide. In Italy, the Italian Sign Language (LIS) was formally acknowledged in May 2021, following an extended period of research highlighting its linguistic and socio-semiotic complexity [3, 29].

In medical settings, communication barriers can lead to misunderstandings and hinder timely care for deaf patients. Employing SL interpreters remains the ideal solution; however, interpreters may not always be available, especially in emergency or resource-limited contexts. Efforts to develop automatic sign language recognition (SLR) technologies have thus gained momentum [10], with the goal of bridging the communication gap between patients and medical staff. Yet privacy concerns persist when deploying visual sensors in sensitive environments like hospitals, where camera-based approaches may be restricted by regulations or ethical constraints [16, 19].

Against this backdrop, radio-frequency (RF) sensing, particularly RADAR, has emerged as a promising avenue for privacy-preserving SLR. RADAR sensors capture motion and velocity information while inherently obscuring fine-grained visual details that could compromise patient identity. This attribute is especially beneficial in dynamic healthcare scenarios, where consistent lighting and uniform backgrounds are not guaranteed [11, 17]. Nevertheless, the existing RADAR-based SLR datasets are comparatively small, often limited to a narrow range of gestures and signers, which limits the generalizability of such systems [4, 5, 21].

This paper thus introduces a privacy-preserving method for LIS recognition in healthcare contexts, leveraging millimeter-wave RADAR technology. Our work focuses on 126 isolated signs relevant to patient-doctor communication, including 100 lexical items and 26 letters of the LIS alphabet. The methodology and the dataset collection strategy leverage RADAR sensing as a privacy-preserving alternative to traditional RGB and depth cameras, offering a lightweight solution that is less sensitive to lighting conditions and environmental occlusions.

2 Related Work

Traditional SLR approaches often rely on specialized hardware—such as sensor-equipped gloves—to capture detailed hand movements [27]. Although these systems provide precise measurements, they can disrupt the natural signing process and omit crucial non-manual cues (e.g., facial expressions). Vision-based solutions have thus become prominent, as they can track both manual and non-manual components in real time using RGB or depth cameras [7, 18]. However, deploying cameras in clinical settings introduces considerable privacy concerns, and variable backgrounds or low-light conditions can further degrade performance [16, 19].

In response, emerging RADAR-based solutions have demonstrated remarkable resilience to lighting and background fluctuations [11, 17]. Micro-Doppler signatures, for instance, capture subtle kinematic patterns of the signer, thereby enabling gesture identification without requiring identifiable visual information [13,25]. Nevertheless, most RADAR-oriented studies have concentrated on American Sign Language (ASL) or generic gestures, featuring limited lexicons and modest datasets [1,19]. This gap underscores the need for broader, more diverse data collections that support robust deep learning models in recognizing complex sign repertoires across various sign languages, including LIS.

Recent efforts to gather multimodal data — incorporating RGB, depth, and skeletal information — reflect the complexity of SL's simultaneous manual and non-manual elements, and require original solutions to perform data integration. For instance, De Coster et al. [8] proposed a Video Transformer Network (VTN) that combines body pose flow and hand-crop features from RGB frames, achieving competitive performance on the AUTSL dataset. Vahdani et al. [28] adopted a multi-stream 3D CNN approach fusing RGB-D, motion, and skeleton data for real-time American Sign Language (ASL) recognition, reporting over 92% accuracy on a newly collected ASL-100-RGBD dataset. However, privacy concerns remain unresolved if visual data are processed off-site or stored without strict protocols.

Integrating RADAR sensing into multimodal datasets can mitigate such constraints by reducing exposure of patient identities, while still retaining essential motion cues [23]. Jhaung et al. [14], Debnath et al. [9], and Arab et al. [2] explored Doppler and FMCW radars for non-contact gesture recognition, demonstrating the viability of radar systems under diverse applications. However, RADAR-based works generally focus on broad activity or hand-gesture classification, rather than capturing the medical-domain lexicon of a sign language. Our approach builds upon these insights by leveraging high-resolution radar data to form a privacy-preserving method tailored to SL patient-doctor communication.

3 Method

3.1 Dataset Description

LIS encompasses both manual (e.g., handshape, orientation, movement) and non-manual (facial expressions, torso, mouth actions) components [3,29]. To address patient-doctor interactions, a dataset was compiled featuring 126 LIS items (100 health-related signs plus 26 letters of the alphabet), drawing on previous LIS corpora guidelines, frequency vocabularies for spoken Italian [6], and incorporating medical and deaf-community insights [5,21]. Following the principles of portability and non-invasiveness outlined in [4], multiple data modalities were simultaneously collected: 13-fps three-antennas RADAR time-domain log-scale data (with an Infineon XENSIV BGT60TR13C 60 GHz sensor); 720p 30-fps RGB-D videos with face tracking points (with an Intel Realsense D455 camera) and additional facial-expression data (through a Microsoft Kinect v1); and 25-fps 1080p depth images/points clouds (using a Stereolabs Zed 2 camera). A single subject performed 205 repetitions per sign, totaling 25,830 sign instances [5,21].

Tab. 1 shows a comparative analysis of our dataset, with respect to existing datasets for SLR. Compared to prior RADAR data collections, our dataset

Table 1. Summary of existing datasets, in terms of data modality, number of signs, number of subjects and number of repetitions per sign. Notation: for RADAR-only data modality, we report the sensor frequency; for "signs", we report by default the number of sign-language signs, as well as letters ("L") or non-sign-language gestures ("G"). * Plus RGB-D and face tracking data.

Dataset	**Modality**	**# signs**	**# subj.**	**# repet.**
SpreadTheSign [13]	RGB	281,672	Unspecified	1
Li et al., [18]	RGB	2,000	3	119
Sincan et al. [26]	RGB-D	226	4	43
Ravi et al. [24]	RGB-D	200	10	10
Jing et sl. [15]	RGB-D	100	42	15
Hassan et al. [12]	RGB-D	10	2	22
Li et al. [17]	8.5 GHz	10	10	10
Lu et al. [19]	24 GHz	5	1	220
McCleary et al. [20]	24 GHz	4	5	250
Park et al. [22]	33 GHz	5L + 6G	10	82
Gurbuz et al. [11]	77 GHz	20	3	3
Wang et al. [30]	77 GHz	6G	1	200
Ours [21]	60 GHz*	100 + 26L	1	205

encompasses both a larger number of symbols and repetitions, yielding a significantly larger dataset and a generally more complex classification task. Some RGB and RGB-D datasets (in particular, SpreadTheSign [13] and Li et al. [18]) feature a comparable or larger number of words, but a significantly lower number of repetitions per word, which hinders their suitability for supervised learning. Additionally, in practical applications, they are subject to the exposure of privacy-sensitive information.

3.2 Data Preprocessing

Our 60 GHz RADAR sensor comprises one transmit antenna and three receive antennas (RX), each recording time-domain signals in logarithmic scale. The radar is configured with a 1 MHz sampling rate, a transmit power level of 31, and an IF gain of 40 dB. Under these settings, the radar achieves a range resolution of 0.0312 m, a maximum range of about 1.60 m, a maximum speed of 4.11 m/s, and a speed resolution of 0.0321 m/s. Each frame is acquired with a repetition time of 0.077 s and a center frequency of 60.50 GHz, ensuring fine-grained motion capture suitable for sign language gestures. Radar data were preprocessed through range and Doppler FFTs with zero-padding, windowing (Blackman-Harris and Chebyshev for range and Doppler, respectively), mean removal, and a spectral threshold of approximately −90 dB (adjustable between −75 dB and −120 dB). An optional Moving Target Indication (MTI) filter further reduces static background signals by zeroing out bins below a certain velocity threshold. The resulting maps (both RDM and RDM-MTI) have a resolution of

128×1024. See the supplementary video for five examples of recorded samples with their correspondin RGB, depth, RDM, and RDM-MTI visualizations.

Signal synchronization was carried out across all sensors (RADAR, RGB-D, depth/point clouds, face features) to ensure time-aligned frames across the 25,830 sign instances, each typically spanning 1–4 seconds. In total, over 6 million radar maps (between RDM/RDM-MTI) are included, providing refined high-resolution range and velocity features for subsequent autoencoder and transformer processing. Please note that our approach only employs RADAR time data from the collected dataset: additional modalities are employed for comparison with methods from the state of the art, or kept for future studies.

3.3 Model Architecture

We address the sign language recognition problem as a classification task. Given the nature of RADAR data, we deal with video-like sequences, where each frame represents an RDM or RDM-MTI map, thus requiring the model to capture both spatial and temporal dynamics for effective classification.

Addressing this challenge directly with a single deep network can lead to excessive model complexity and computational costs, particularly given the high dimensionality and sequential nature of radar data. To mitigate these issues, our approach decouples the problem into two distinct stages: frame compression through an autoencoder and sequence classification with a transformer network.

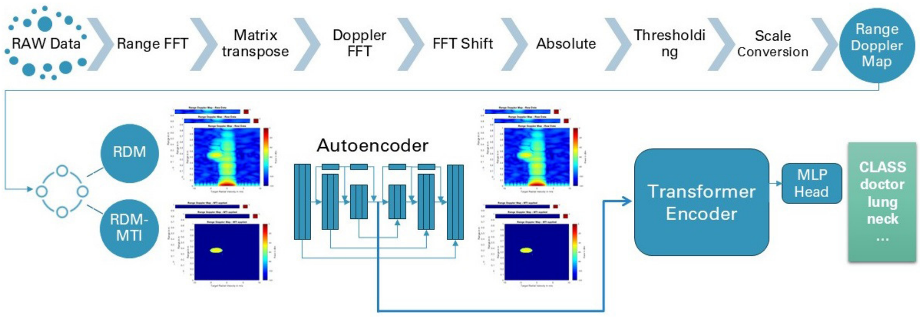

Fig. 1. Overview of the end-to-end architecture of the proposed method.

Autoencoder for RADAR feature extraction. The proposed autoencoder network is designed to learn a compact and meaningful representation of radar data for sign language recognition. The fundamental objective of this autoencoder is to reconstruct the original input RADAR map, either RDM or RDM-MTI, while distilling its essential features into a low-dimensional embedding, which is used later as input to the transformer classifier. Due to the significant differences in nature between RDM/RDM-MTI maps and natural images, we do not employ pretrained models; we thus design a custom CNN architecture[5], with symmetric encoder and decoder, each featuring 9 convolutional layers (grouped into three blocks, for each feature map resolution, before downsampling). We add residual connnections between the input and output of each block, to ease gradient flow. The bottleneck of the model yields a 256-dimensional compressed representation of the input signal.

Transformer classifier. Following the autoencoder pretraining, the classification stage of the proposed approach is based on a transformer network, which aims to effectively classify the sign language representations captured in the 256-dimensional embeddings. The objective of this transformer-based classifier is to leverage its powerful sequence modeling capabilities to capture patterns in the RADAR signal, corresponding to the dynamic gestures of sign language. The transformer architecture is particularly suited for this task due to its inherent ability to model long-range dependencies and capture the sequential nature of sign language gestures. Moreover, it naturally handles variable-length sequences, which is necessary in our scenario due to the different lengths of the recorded sign, ranging from 13 to 66 RDM/RDM-MTI frames.

Our approach utilizes a custom transformer design that processes the output embeddings from the autoencoder. The autoencoder 1024-dimensional embeddings are first projected on a 64-dimensional space; in order to support the combination of multiple RADAR modalities with the same model, whenever multiple modalities are employed we fuse the autoencoder embeddings via convolutional layers by treating them as channels, before feeding them to the initial projection layer. We then augment the sequence with a learnable class token and positional embeddings to maintain temporal context, before being fed to the transformer. The Transformer architecture employs six transformer layers, with standard multi-head self-attention and eight attention heads per layer. At the output layer of the transformer, we feed the resulting class token embedding into a linear classification layer, estimating class logits for the dataset's 126 classes and minimizing a standard cross-entropy loss.

3.4 Training procedure

We partition the dataset into 60% training, 20% validation, and 20% test sets. No data augmentation is carried out. As mentioned above, we preliminarily train the autoencoder model, in order to learn a good representation of RADAR signals at a single time step, to simplify the following sequence classification task. We train the model for 15 epochs with a batch size of 16, adopting the AdamW optimizer with $\beta_1 = 0.9, \beta_2 = 0.999$ and a learning rate of $5 \cdot 10^{-5}$.

The autoencoder is then frozen and used only to extract the embeddings from the dataset samples. We train the transformer for 700 epochs, with a batch size of 256 and a learning rate of 10^{-4}, with an AdamW optimizer as above. Gradient clipping is enforced at a maximum norm of 5, and a weight decay of $5 \cdot 10^{-6}$ is

[5] Please refer to the public source code for model details.

applied to mitigate overfitting. All experiments are performed in `float16` mixed precision using 5×NVIDIA A100 40GB and 4×NVIDIA A6000 Ada 48GB. This two-stage process alleviates GPU memory constraints by avoiding full end-to-end training on raw data for every epoch.

The above hyperparameters are chosen based on the model's performance on the validation set.

4 Experimental Results

In this section, we present a thorough performance assessment of our approach, comparing our results to state-of-the-art methods and justifying architectural choices described above. We report results in terms of classification accuracy, sensitivity, specificity, and F1-score.

Table 2 reports the comparison, on our dataset, of the proposed approach with state-of-the-art solutions, drawn from both RGB/RGB-D approaches and radar-based methods. Where the source code is not provided by the authors, we wrote our own implementations, adhering to the reference description at the best of our possibilities. Model selection is performed based on validation performance, using default hyperparameters. For our approach, we report results corresponding to different input modalities (single/multi-antenna, RDM only, RDM-MTI only and the combination of RDM and RDM-MTI).

Table 2. Comparison with state-of-the-art methods in sign language recognition. The "Modality" column indicates the primary sensor modality employed by the method.

Method	**Modality**	**Acc.**	**Sens.**	**Spec.**	**F1-score**
De Coster et al. [8]	RGB	88.4 ± 3.7	78.6 ± 3.4	98.2 ± 3.5	79.7 ± 3.9
Vahdani [28]	RGB-D	84.1 ± 2.9	83.1 ± 4.1	85.1 ± 2.1	84.2 ± 2.2
ResNet(2+1)d	RGB-D	74.6 ± 3.9	66.4 ± 3.1	82.8 ± 3.0	59.9 ± 3.3
Jhaung et al. [14]	RADAR	71.9 ± 3.3	61.9 ± 2.3	81.9 ± 3.4	69.2 ± 2.9
Debnath et al. [9]	RADAR	79.3 ± 3.5	83.3 ± 3.3	75.3 ± 3.1	79.3 ± 3.9
Arab et al. [2]	RADAR	81.0 ± 3.4	65.9 ± 3.1	96.1 ± 2.8	29.6 ± 2.9
Ours	RDM	88.3 ± 1.1	79.9 ± 0.8	96.7 ± 0.9	86.6 ± 1.2
	3×RDM	91.7 ± 0.6	86.9 ± 0.6	96.5 ± 0.7	87.5 ± 0.5
	MTI	84.9 ± 0.9	73.9 ± 0.7	95.9 ± 0.8	80.0 ± 0.8
	3×MTI	86.1 ± 0.6	77.6 ± 0.5	94.6 ± 0.5	83.3 ± 0.7
	RDM + MTI	91.4 ± 0.6	84.1 ± 0.5	98.7 ± 0.5	86.9 ± 0.5
	3×RDM + 3×MTI	93.6 ± 0.5	87.9 ± 0.5	99.3 ± 0.6	87.7 ± 0.5

Our approach achieves the best performance when employing three antenna streams and both RDM and RDM-MTI channels, reflecting the advantage of integrating multiple radar perspectives and emphasizing moving-target information. Notably, even though radar systems typically exhibit lower performance compared to RGB-based systems in prior work, our solution achieves comparable, if not better results on several metrics, demonstrating the viability of privacy-preserving, radar-centric solutions for sign language recognition.

We also evaluate the effectiveness of the proposed autoencoding approach for feature extraction, across two different aspects. First, we compare the proposed autoencoder backbone to other popular architectures, i.e., AlexNet and ResNet-18. We do not test more powerful architectures, due to the relative simplicity and lack of complex and varying patterns of RDM images. Second, we examine the effect of fine-tuning the autoencoder while training the transformer classifier (we remind that, in our default configuration, the autoencoder is frozen during transformer training). Tab. 3 shows the results of this analysis. In all cases, keeping the autoencoder frozen improves performance, due to an increased overfitting while training the transformer (as also empirically observed in the training and validation loss curves). Moreover, the proposed autoencoder architecture achieves higher performance than other networks, which may be explained, as mentioned, to the lack of complex features in the input data, enabling a simpler model to better fit the data distribution.

Table 3. Comparison of autoencoder architectures and training strategies.

Backbone	**Acc.**	**Sens.**	**Spec.**	**F1-score**
AlexNet	61.4 ± 1.6	43.9 ± 1.9	78.9 ± 1.6	54.1 ± 1.3
↪ Frozen	73.9 ± 1.7	68.8 ± 2.0	79.0 ± 2.1	66.6 ± 1.8
ResNet-18	84.3 ± 0.4	76.2 ± 0.4	92.4 ± 0.6	78.2 ± 0.6
↪ Frozen	88.1 ± 1.2	81.7 ± 0.6	94.5 ± 0.9	79.9 ± 0.5
Ours	91.9 ± 0.6	86.1 ± 0.5	97.4 ± 0.6	85.5 ± 0.7
↪ **Frozen**	93.6 ± 0.5	87.9 ± 0.5	99.3 ± 0.6	87.7 ± 0.5

5 Conclusion

In this paper, we have presented a privacy-preserving, RADAR-based solution for SLR in a medical context, using a multimodal dataset of 126 signed items (100 medically relevant terms and 26 letters). Our two-stage pipeline —involving a convolutional autoencoder backbone followed by a Transformer classifier — demonstrates state-of-the-art performance, underlining the feasibility of non-visual sensing methods for sign language interpretation in sensitive healthcare environments. In the future, we intend to integrate face and lip-joint tracking into the model pipeline, allowing the system to disambiguate visually similar signs based on subtle mouth movements. Such enhancements can prove especially valuable when indicating specific body parts or health conditions that share comparable manual gestures; moreover, techniques such as landmark extraction would maintain the same privacy-preserving properties as the pure RADAR-base system.

Acknowledgments. F. Proietto Salanitri, I. Kavasidis and C. Spampinato acknowledge financial support from PNRR MUR project PE0000013-FAIR. The work of Simone Palazzo has been supported by MUR in the framework of PNRR Mission 4, Component 2, Investment 1.1, PRIN, under project RESILIENT, CUP E53D23016360001. R. Mineo is a PhD student enrolled in the XXXVII cycle of National PhD in Artificial Intelligence, course on Health and life sciences, organized by University Campus Bio-Medico of Rome.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Adeoluwa, O.O., Kearney, S.J., Kurtoglu, E., Connors, C.J., Gurbuz, S.Z.: Near real-time asl recognition using a millimeter wave radar. In: Radar Sensor Technology XXV. vol. 11742, pp. 173–184. SPIE (2021)
2. Arab, H., Ghaffari, I., Chioukh, L., Tatu, S.O., Dufour, S.: A convolutional neural network for human motion recognition and classification using a millimeter-wave doppler radar. IEEE Sensors Journal **22**(5), 4494–4502 (2022)
3. Branchini, C., Mantovan, L., et al.: A Grammar of Italian Sign Language (LIS). Edizioni Ca'Foscari (2020)
4. Caligiore, G.: Codifying the Body: Exploring the Cognitive and Socio-semiotic Framework in Building a Multimodal Italian Sign Language (LIS) Dataset [Ph. D. thesis, University of Catania]. Ph.D. thesis, University of Catania (2024)
5. Caligiore, G., Mineo, R., Spampinato, C., Ragonese, E., Palazzo, S., Fontana, S.: Multisource approaches to italian sign language (lis) recognition: Insights from the multimedalis dataset. In: 2024 Tenth Italian Conference on Computational Linguistics (CLiC-it) (2024)
6. Cecchetto, C., Giudice, S., Mereghetti, E., et al.: La raccolta del corpus lis. In: Grammatica, lessico e dimensioni di variazione nella LIS,, pp. 55–67. Franco Angeli (2011)
7. Chen, T., Dong, X., Chen, Y.: Gesture recognition with feature fusion using fmcw radar. In: Seventh Asia Pacific Conference on Optics Manufacture and 2021 International Forum of Young Scientists on Advanced Optical Manufacturing (APCOM and YSAOM 2021). vol. 12166, pp. 353–362. SPIE (2022)
8. De Coster, M., Van Herreweghe, M., Dambre, J.: Isolated sign recognition from rgb video using pose flow and self-attention. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3441–3450 (2021)
9. Debnath, B., Ebu, I.A., Biswas, S., Gurbuz, A.C., Ball, J.E.: Fmcw radar range profile and micro-doppler signature fusion for improved traffic signaling motion classification. In: 2024 IEEE Radar Conference (RadarConf24). pp. 1–6. IEEE (2024)
10. Fontana, S., Caligiore, G.: Italian sign language (lis) and natural language processing: an overview. NL4AI@ AI* IA (2021)
11. Gurbuz, S.Z., Gurbuz, A.C., Malaia, E.A., Griffin, D.J., Crawford, C., Rahman, M.M., Aksu, R., Kurtoglu, E., Mdrafi, R., Anbuselvam, A., et al.: A linguistic perspective on radar micro-doppler analysis of american sign language. In: 2020 IEEE international radar conference (RADAR). pp. 232–237. IEEE (2020)

12. Hassan, S., Berke, L., Vahdani, E., Jing, L., Tian, Y., Huenerfauth, M.: An isolated-signing rgbd dataset of 100 american sign language signs produced by fluent asl signers. In: Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives. pp. 89–94 (2020)
13. Hilzensauer, M., Krammer, K.: A multilingual dictionary for sign languages:" spreadthesign". In: ICERI2015 Proceedings. pp. 7826–7834. IATED (2015)
14. Jhaung, Y.C., Lin, Y.M., Zha, C., Leu, J.S., Köppen, M.: Implementing a hand gesture recognition system based on range-doppler map. Sensors **22**(11), 4260 (2022)
15. Jing, L., Vahdani, E., Huenerfauth, M., Tian, Y.: Recognizing american sign language manual signs from rgb-d videos. arXiv preprint arXiv:1906.02851 (2019)
16. Kulhandjian, H., Sharma, P., Kulhandjian, M., D'Amours, C.: Sign language gesture recognition using doppler radar and deep learning. In: 2019 IEEE globecom workshops (GC Wkshps). pp. 1–6. IEEE (2019)
17. Li, B., Yang, J., Yang, Y., Li, C., Zhang, Y.: Sign language/gesture recognition based on cumulative distribution density features using uwb radar. IEEE transactions on instrumentation and measurement **70**, 1–13 (2021)
18. Li, D., Rodriguez, C., Yu, X., Li, H.: Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 1459–1469 (2020)
19. Lu, Y., Lang, Y.: Sign language recognition with cw radar and machine learning. In: 2020 21st International Radar Symposium (IRS). pp. 31–34. IEEE (2020)
20. McCleary, J., García, L.P., Ilioudis, C., Clemente, C.: Sign language recognition using micro-doppler and explainable deep learning. In: 2021 IEEE Radar Conference (RadarConf21). pp. 1–6. IEEE (2021)
21. Mineo, R., Caligiore, G., Spampinato, C., Fontana, S., Palazzo, S., Ragonese, E.: Sign language recognition for patient-doctor communication: A multimedia/multimodal dataset. In: 2024 IEEE 8th Forum on Research and Technologies for Society and Industry Innovation (RTSI). pp. 202–207. IEEE (2024)
22. Park, J., Jang, J., Lee, G., Koh, H., Kim, C., Kim, T.W.: A time domain artificial intelligence radar for hand gesture recognition using 33-ghz direct sampling. In: 2019 Symposium on VLSI Circuits. pp. C24–C25. IEEE (2019)
23. Rahman, M.M., Mdrafi, R., Gurbuz, A.C., Malaia, E., Crawford, C., Griffin, D., Gurbuz, S.Z.: Word-level sign language recognition using linguistic adaptation of 77 ghz fmcw radar data. In: 2021 IEEE Radar Conference (RadarConf21). pp. 1–6. IEEE (2021)
24. Ravi, S., Suman, M., Kishore, P., Kumar, K., Kumar, A., et al.: Multi modal spatio temporal co-trained cnns with single modal testing on rgb–d based sign language gesture recognition. Journal of Computer Languages **52**, 88–102 (2019)
25. Ren, A., Wang, Y., Yang, X., Zhou, M.: A dynamic continuous hand gesture detection and recognition method with fmcw radar. In: 2020 IEEE/CIC International Conference on Communications in China (ICCC). pp. 1208–1213. IEEE (2020)
26. Sincan, O.M., Keles, H.Y.: Autsl: A large scale multi-modal turkish sign language dataset and baseline methods. IEEE access **8**, 181340–181355 (2020)
27. Stokoe Jr, W.C.: Sign language structure: An outline of the visual communication systems of the american deaf. Journal of deaf studies and deaf education **10**(1), 3–37 (2005)

28. Vahdani, E., Jing, L., Huenerfauth, M., Tian, Y.: Multi-modal multi-channel american sign language recognition. International Journal of Artificial Intelligence and Robotics Research **1**(01), 2450001 (2024)
29. Volterra, V., Roccaforte, M., Di Renzo, A., Fontana, S.: Italian Sign Language from a cognitive and socio-semiotic perspective: Implications for a general language theory. John Benjamins (2022)
30. Wang, Y., Ren, A., Zhou, M., Wang, W., Yang, X.: A novel detection and recognition method for continuous hand gesture using fmcw radar. IEEE Access **8**, 167264–167275 (2020)