Radar-Based CNN Recognition of Gesture Primitives for SignWriting Educational Feedback

Fatma Makhlouf
Higher School Communication of Tunis, SUP'COM
University of Carthage
Tunisia
Fatma.Makhlouf@supcom.tn

Fethi Mejri
National Engineering School of Tunis
University of El Manar
Tunisia
fethi.mejri.syscom@gmail.com

Neji Youssef
SUP'COM
University of Carthage
Tunisia
neji.youssef@supcom.tn

***Abstract*—Sign languages depend on organized combinations of hand shapes and movements, which are often represented through formal writing systems, such as SignWriting. Although most sign language recognition systems depend on vision-based sensors, these approaches raise privacy concerns and require controlled environments. This paper presents a pilot study investigating the feasibility of recognizing SignWriting-inspired gesture primitives using a Doppler radar sensor and convolutional neural networks (CNNs). Using uRAD radar measurements, in-phase and quadrature (I/Q) signals were collected for a set of fundamental gesture primitives reflecting static holds and directional hand motions. A CNN was trained to classify these primitives from radar time–range representations. The experimental results show that the classification accuracy exceeds the random baseline performance, demonstrating that the Doppler radar captures discriminative motion patterns relevant to symbolic sign representations. Although performance is constrained by dataset size and sensor resolution, the proposed approach establishes a foundation for radar-based sign-writing systems and assistive technologies for deaf and mute users.**

I. Introduction

Communication is the primary means of interaction between people, especially for those who are deaf or mute. Without the ability to express themselves, these individuals may struggle with feelings of hopelessness, as they often depend on gestures to convey their needs to others. The expressions and gestures used by mute individuals are vital for communicating their requirements to their caregivers. Therefore, developing an application or system tailored to these communities could significantly enhance their quality of life and prevent feelings of isolation.

In the next section, we discuss the limitations that led to this finding.

II. Challenges and Gaps Identified

Numerous research projects have been conducted to provide technological support, such as developing IoT and cloud-based solutions, such as apps for translating sign language to help people communicate better. We will discover and analyze the strengths and limitations of these studies.

A. Tunisian Sign Language

This study [1] built a dataset of 2,000 images containing 12 two-handed asymmetrical Tunisian signs and applied transfer learning to classify static signs, reporting 98.29% accuracy. The second work [2] provides a new TSL dataset and examines advanced AI approaches for video-based TSL recognition, enabling further experimental work on the Tunisian data. This study [4] describes TunSLR-25, a static Tunisian Sign Language recognition system developed using deep learning techniques. This study [5] investigated vision-based techniques for an ecological (real-world) system that recognizes Tunisian sign language, emphasizing robustness in uncontrolled settings. [6] This article focuses on a practical report covering deep learning, computer vision, and mobile development for real-time recognition of Tunisian sign language, as described in the supplied metadata. This study proposes a two-stage pipeline that leverages 2D joint locations to improve sign recognition performance, which is relevant to TSL research focusing on facial and body cues [9]. This manuscript describes a system for recognizing sign language that combines computer vision, machine learning, and signal processing, with an emphasis on real-time implementation considerations [10].

B. Challenges

While Tunisian Sign Language (TSL) is used, challenges include limited education, employment, and social inclusion, although associations are growing. Additionally, there are numerous barriers; the insufficient number of people who use sign language complicates access to services for deaf individuals on a daily basis, according to the Tunisian National Deaf Center.

Research [1], [7] overlooks the limitations of other sign languages. For example, images and videos may violate the privacy of humans, especially deaf or mute people, who are naturally sensitive to such issues. Forcing them to face a camera is disrespectful and intrusive. The use of video or image capture sensors can further exacerbate this sensitivity. Additionally, these methods do not safeguard individuals' personal information during the preprocessing and processing stages.

However, the limitations of using video and camera sensors, including lighting bias and privacy concerns, have been identified. In contrast, radar sensors have numerous advantages. Radar can function effectively under low-light conditions, navigate through fog, and penetrate clutter. These findings suggest that this study is relevant to real-world applications, such as healthcare monitoring. This study bridges radar sensing and symbolic gesture classification, targeting assistive applications in which robustness, privacy, and simplicity are essential.

III. The Proposed Methodology

A. Materials and system architecture

1) Materials: This study utilized a frequency-modulated continuous wave (FMCW) radar sensor, known as the uRAD radar, which operates at a frequency of 24 GHz. The uRAD radar can function as a portable radar system, providing various modes that enhance its ability to adapt dynamically to different detection scenarios. See fig. 3.

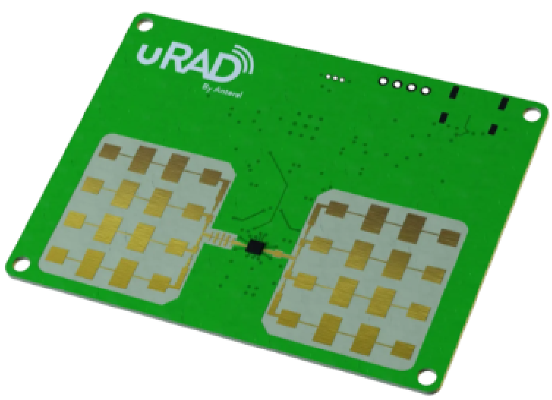

Fig. 1. The uRAD V1.2

FMCW radar technology operates by continuously transmitting frequency-modulated waves, typically employing linear frequency modulation (LFM). By computing the frequency shift (beat frequency) between the emitted and reflected signals, the FMCW radar extracts critical information regarding the distance and velocity of detected objects. This continuous frequency variation provides two fundamental advantages: superior spatial resolution, which enables the precise discrimination of proximal and high-speed targets, and robust interference immunity, which mitigates the impact of ambient electromagnetic disturbances and ensures stable performance in complex environments [19].

The uRAD USB v1.2 radar supports four distinct operational modes (Table 2), enabling dynamic adaptability to various detection scenarios. These features collectively position radar as an efficient and scalable solution for motion sensing, autonomous navigation, and security. The following table presents the waveform representations corresponding to the four operational modes of the uRAD USB v1.2 radar, providing a comparative visualization of its signal dynamics.

The waveform generated by the FMCW radar is known as a chirp signal. In linear frequency modulation (LFM), the frequency varies in an ascending (up-chirp) or descending (down-chirp) manner, with the extent of this variation defined as the signal bandwidth [20]. A widely adopted approach involves combining up-chirp and down-chirp sequences to form a triangular modulation pattern. Figure 2 presents the chirp signal model generated by the FMCW radar, emphasizing its essential parameters.

The frequency of the emitted signal over time can be represented as follows: Where b = B T is referred to as the frequency slope, T represents the duration of the chirp, B indicates the frequency bandwidth traversed, $t' \in [0, T], \quad k = 0, \ldots, K-1$

Therefore, the signal emitted by the FMCW radar can be modeled as:

$$s(t) = \cos\left(2\pi f_C t + \phi(t)\right) \tag{1}$$

where time t and phase j(t) are given as follows:

$$t = kT + t' \tag{2}$$

$$\phi(t) = p_k b T^2 + p_b t'^2 \tag{3}$$

The relationship between the signal reflected from an object and the object can be expressed as follows:

$$y(t) \approx e^{-j2\pi f_B t'} \tag{4}$$

$$e^{-j2\pi f_D kT} \tag{5}$$

where # signifies the complex coefficient dependent on gain, fB indicates the beat frequency, and fD represents the Doppler frequency. The radar's receiving pathway is responsible for multiplying the received signal with a replica of the transmitted signal and filtering it out. Finally, the speed and distance to the object can be determined by knowing the values of fB and fD [21], where the speed of light is indicated by c.

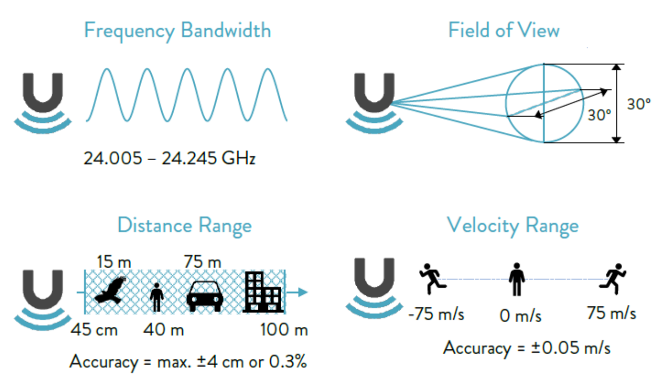

Fig. 2. The main features of uRAD

2) System design: The proposed system aims to classify static hand gestures captured by a low-cost Doppler radar sensor into symbolic representations suitable for assistive and educational applications for deaf and mute users to facilitate communication. The overall pipeline consists of radar data acquisition, signal preprocessing, tensor construction, convolutional neural network (CNN) training, and gesture classification.

The input to the system is the raw in-phase (I) and quadrature (Q) radar baseband signals, whereas the output is a discrete gesture class corresponding to a predefined symbol.

B. Experimental procedures

1) Load Configuration of uRAD: The radar module supports USB connectivity, enabling seamless integration with computational platforms for advanced data processing and controls. It features 16 transmit and 16 receive antennas, providing a 30° field-of-view (FoV) in both the horizontal and vertical planes. Powered by a 5V USB supply, the system was engineered to operate reliably under extreme environmental conditions, maintaining performance within a temperature range of 20–65°C.

TABLE I
URAD SENSOR CONFIGURATION

Parameter	Value
Mode	FMCW
Number of Samples per Cycle	200
Sampling Rate	24 Hz
Recording Duration per Gesture	180 seconds

2) Data Acquisition: We used a uRAD USB Doppler radar sensor to record the hand movement data. The radar operates in continuous-wave mode and sends out baseband in-phase (I) and quadrature (Q) signals. The data were collected at a rate of approximately 100 Hz.

Every time a gesture occurred, it was recorded over a set time of 300 frames (approximately 3 s). To ensure that all recordings were the same, the radar was set up at a fixed distance and angle from the subject.

Three different hand gestures were defined, and each represented a different symbolic class. These gestures were selected to demonstrate various motion orientations and dynamics, facilitating a distinct separation in the radar signal space.

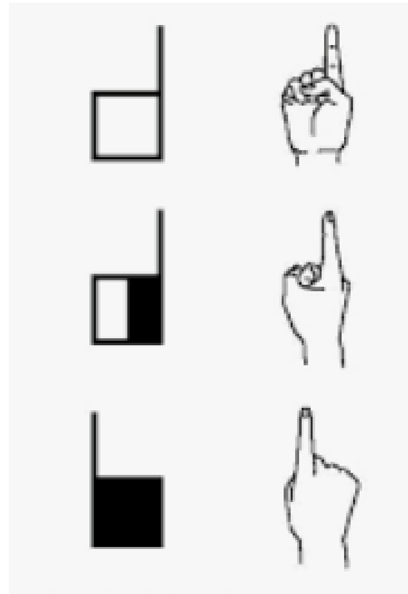

Fig. 3. The gestures collected

3) Neural Network Model: A convolutional neural network (CNN) is an initial deep learning architecture characterized by a series of layers. CNN is an algorithm derived from the Multilayer Perceptron (MLP), which is intended for processing 2D data. The CNN technique has been extensively utilized for image data and is categorized as a deep neural network. The distinction between MLP and CNN lies solely in the dimensionality of the neurons; MLP utilizes one-dimensional neurons, whereas CNN employs two-dimensional neurons. This study employed three CNN situations, each characterized by a distinct number of layers, with the objective of identifying the optimal model. CNN will categorize five courses that utilize the 5-letter ASL gesture. The detection system was derived from a radar, and a programming algorithm was established within the system flowchart.

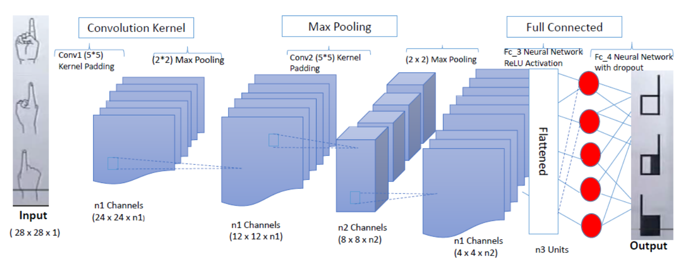

Fig. 4. The Architecture of CNN

A CNN has four key components:

KERNEL OR FILTER OR FEATURE DETECTORS.

In CNN, a kernel is employed to extract features from images.

$$[i - k] + 1 i\text{: } \textit{Size of input , K: Size of kernel}$$

STRIDE.

Stride is a parameter of the neural network filter that alters the extent of movement across a picture or video. We commenced with stride 1; hence, it will proceed sequentially. Utilizing a stride of 2 results in the selection of values by omitting the subsequent two pixels.

$$\left\lfloor \frac{i-k}{s} \right\rfloor + 1$$

i: Size of input , K: Size of kernel, S: Stride (7)

PADDING.

Padding pertains to convolutional neural networks, denoting the quantity of pixels included in an image during processing by the CNN kernel. If the padding in a CNN is configured to zero, every new pixel value will become zero. Utilizing a filter or kernel to process an image results in a reduction in its dimensions. We must prevent this to maintain the original dimensions of the image for the extraction of low-level characteristics. Consequently, we incorporated more pixels outside the image boundaries.

$$\left\lfloor \frac{i-k+2p}{s} \right\rfloor + 1$$

where i: Size of input , K: Size of kernel, S: Stride, p :Padding

POOLING.

Pooling in convolutional neural networks is a method for generalizing features obtained via convolutional filters, aiding the network in recognizing features irrespective of their spatial placement within the image.

FLATTEN.

Flattening is used to convert all the resultant 2-dimensional arrays from the pooled feature maps into a single long continuous linear vector. The flattened matrix was fed as input to the fully connected layer to classify the image.

LAYERS USED TO BUILD CNN.

Convolutional Neural Networks (CNNs) are distinguished from other neural networks by their superior performance with image, speech, and audio signal inputs. They have three main types of layers, which are

nosep

- `Convolutional layer`
- `Pooling layer`
- `Fully-connected (FC) layer`

Convolutional layers This is the

The first component of a Convolutional Neural Network (CNN). The major mathematical activity that this program performs is called convolution, which is the use of a sliding window function on a matrix of pixels to represent an image. The matrix's sliding function is called a "kernel" or "filter." In the convolution layer, several filters of the same size are used to find certain patterns in the image, such as the way the digits curve, the edges, and the complete shape of the digits.

We employed small grids that moved across the image in the convolution layer. These grids are referred to as filters or kernels. Each small grid functions as a tiny magnifying glass that searches for specific features, lines, or curves in the image. As the small grid advances across the photo, it creates a new grid that indicates where the patterns are detected.

The convolutional layer extracts discriminative spatiotemporal features by applying learnable kernels across the multichannel radar input, followed by a ReLU activation to introduce nonlinearity.

Each convolutional layer applies a set of learnable filters to the multi-channel radar input. For each output feature map k, the convolution operation aggregates local spatio-temporal information across both in-phase and quadrature channels, followed by a ReLU non-linearity to enhance discriminative feature learning.

$$\frac{W-F+2P}{S} + 1 \tag{9}$$

Rectified Linear Unit (ReLU for short) — A ReLU activation function is utilized after each convolution step. This function helps the network understand the complex relationships between image features, making it better at recognizing patterns. It also helps alleviate the vanishing gradient issues.

Pooling layers The purpose of the pooling layer is to extract the most salient characteristics from the convoluted matrix. This is done using methods that reduce the size of the feature map (convoluted matrix), which helps to use less memory when training the network. Pooling plays a crucial role in reducing the overfitting. The following primary aggregation functions may be utilized:

nosep

- `Max pooling`
- `Sum pooling`
- `Average pooling`

Max pooling It refers to the maximum value extracted from the feature map.

Sum pooling-refers to the aggregation of all values within the feature map using summation.

Average pooling—represents the mean of all values.

Fully connected layers The final layers of the convolutional neural network utilize inputs from the flattened one-dimensional matrix produced by the last pooling layer. ReLU activation functions are used to introduce nonlinearity.

A softmax prediction layer was employed to produce probability values for each potential output label, with the

TABLE III

COMPARISON WITH RELATED RADAR-BASED GESTURE RECOGNITION STUDIES

Work	Radar Type	Model	Gestures	Reported Accuracy
This Work (Static Hand Gestures)	Doppler (uRAD)	CNN (2-channel I/Q)	3	**100%**
FSK Radar with CNN [12]	FSK Radar	CNN	Not specified	93.67%
TRANS-CNN [13]	mmWave (IWR1642)	CNN + Attention	10+	98.5%
TinyRadarNN [14]	Short-Range Radar	CNN + TCN	11	92.4% (single-user)
Fine-tuned ResNet (MR-HGR) [15]	mmWave Radar	ResNet / ShuffleNet	6	93.47%

final predicted label being the one that attained the highest probability score.

Application. This study employed a different number of layers to categorize the five classes. The radar detection system was processed, and the system flowchart designed a programming procedure.

IV. RESULTS AND DISCUSSION

A. Classification Performance

The trained CNN achieved a 100% classification accuracy on the test dataset. The confusion matrix indicated a flawless distinction among all the gesture classes, with no misclassifications detected.

TABLE II
CONFUSION MATRIX FOR RADAR-BASED GESTURE CLASSIFICATION

True / Predicted	Gesture 1	Gesture 2	Gesture 3
Gesture 1	10	0	0
Gesture 2	0	7	0
Gesture 3	0	0	8

The confusion matrix presented in Table II demonstrates a perfect classification across all gesture classes. No inter-class confusion was observed, indicating that the proposed CNN successfully learned discriminative spatiotemporal features from the radar I/Q data.

B. Interpretation of Results

The findings revealed that radar-based I/Q representations possess sufficient distinctive information to differentiate between symbolic hand gestures. The convolutional filters effectively captured the spatiotemporal motion patterns associated with hand orientation and subtle movements.

The outstanding classification outcomes highlight the efficiency of the proposed CNN architecture in controlled settings and confirm the appropriateness of radar sensing for detecting assistive gestures.

The results revealed that the radar-based CNN effectively discerned the unique motion patterns associated with each gesture. The perfect classification outcomes suggest that the selected gestures possess sufficiently distinct micro-Doppler signatures in a controlled environment.

It is important to note that the dataset size and experimental setup were limited. Therefore, the results should be interpreted as proof of concept rather than a claim of generalization to unconstrained environments.

C. Scientific Validation and Limitations

Although the achieved accuracy is commendable, the dataset size and experimental setup are constrained. The present study was limited to static gestures and a singular acquisition environment. Future research will explore generalization across multiple users, dynamic gestures, and different radar configurations.

V. CONCLUSION

This study presents a system for classifying gesture symbols using radar technology in conjunction with convolutional neural networks. By employing raw I/Q radar data, the method achieves perfect classification accuracy for a set of symbolic hand gestures. The results underscore the potential of integrating radar sensing with neural networks for applications aimed at assisting and educating individuals who are deaf or mute.

ETHICAL CONSIDERATIONS

This study employed non-invasive radar sensing that did not collect visual or personally identifiable information. No biometric identification was involved in this study. The data collection process adhered to ethical guidelines aimed at safeguarding individuals' privacy and ensuring that participants were informed about the procedures.

REFERENCES

[1] E. Daknou, H. Hermessi, and N. Tabbane, "Tunisian Sign Language Recognition System of Static Two-Handed Asymmetrical Signs using Deep Transfer Learning," *MCN*, vol. 9, no. 27, Oct. 2024, doi: 10.54455/mcn2702.

[2] H. Fourati and L. A. Saidane, "Artificial Intelligence Use in Sign Recognition: A Dataset in Tunisian Sign Language," in *Proc. IEEE IWCMC*, May 2025, doi: 10.1109/IWCMC65282.2025.11059445.

[3] P. A. Prabha, M. Daanish, N. Kumar, and N. M. S. Kumaar, "DeepSignSpeak: Deep Learning Based Sign Language Recognition and Regional Language Translation," in *Proc. ACET*, Aug. 2024, doi: 10.1109/ACET61898.2024.10730664.

[4] S. Aloui and F. Zhou, "TunSLR-25: A New Static Tunisian Sign Language Recognition System," in *Lecture Notes in Computer Science*, vol. 14353, Springer, 2024, doi: 10.1007/978-3-031-81806-6_9.

[5] S. Fakhfakh and Y. B. Jemaa, "Towards Ecological Tunisian Sign Language Recognition System," in *Lecture Notes in Computer Science*, pp. 362–371, Springer, 2024, doi: 10.1007/978-3-031-57840-3_33.

[6] R. Jerbi, W. Chakroun, and M. Abdelfatteh, "Mobile Application for Tunisian Sign Language (TunSL) Recognition," Tech. Rep., 2022. [Online]. Available: doi:10.13140/RG.2.2.31806.36162

[7] M. E. Askri, H. Basly, R. Bchir, M. A. Zayene, and F. Sayadi, "Tunisian Sign Language Recognition and Translation Using Deep Learning," in *Lecture Notes in Computer Science*, Springer, 2024, doi: 10.1007/978-3-031-82156-1_3.

[8] S. Fakhfakh and Y. B. Jemaa, "Deep Learning Shape Trajectories for Isolated Word Sign Language Recognition," *International Arab Journal of Information Technology*, vol. 19, no. 4, pp. 660–666, Jan. 2022, doi: 10.34028/iajit/19/4/10.

[9] M. Jebali, L. Trabelsi, H. Harrouch, R. Triki, and S. Mohamed, "Development of Deep Learning-Based Models Highlighting the Significance of Non-Manual Features in Sign Language Recognition," *AIMS Mathematics*, Jan. 2025, doi: 10.3934/math.2025898.

[10] K. K. Sukhadan, V. D. Bakhade, G. S. Thakare, K. G. Dhanbhar, and A. C. Deshmukh, "Sign Language Recognition System," *Int. J. Res. Appl. Sci. Eng. Technol.*, Mar. 2024, doi: 10.22214/IJRASET.2024.58758.

[11] V. C. Chen, F. Li, S. Ho, and H. Wechsler, "Micro-Doppler effect in radar," *IEEE Aerospace and Electronic Systems Magazine*, vol. 21, no. 5, pp. 16–23, 2011.

[12] Y. Kim and H. Ling, "Human activity classification based on micro-Doppler signatures using deep convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5129–5141, 2018. DOI: 10.1109/TGRS.2018.2794598.

[13] P. Molchanov et al., "Classification of hand gestures using radar sensors and convolutional neural networks," in *Proc. IEEE Radar Conference*, 2016. DOI: 10.1109/RADAR.2016.7485199.

[14] Z. Zhang et al., "Radar-based hand gesture recognition using deep learning," *Sensors*, vol. 19, no. 7, 2019. DOI: 10.3390/s19071545.

[15] H. Cooper et al., "SignWriting: A universal system for writing sign languages," *Sign Language Studies*, vol. 11, no. 3, pp. 343–369, 2011. DOI: 10.1353/sls.2011.0003.