Journal of Computer and Communications, 2026, 14(9), 1-9
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Development of a System for the Recognition of Isolated Signs in Text of the Niger Sign Language (LSNi) Based on Neural Networks

Bachir Moussa Idi, Chaibou Kadri, Ibrahim Bouwey Alley, Yahaya Morou Ganda, Harouna Naroua

Département de Mathématiques et Informatique, Faculté des Sciences et Techniques, Université Abdou Moumouni, Niamey, Niger
Email: bachir.moussaidi@yahoo.fr, kadrichaibou73@gmail.com, ibrahim.bouwey.98@gmail.com, y.morouganda.35116@gmail.com, hnaroua@yahoo.com

How to cite this paper: Moussa Idi, B., Kadri, C., Bouwey Alley, I., Morou Ganda, Y. and Naroua, H. (2026) Development of a System for the Recognition of Isolated Signs in Text of the Niger Sign Language (LSNi) Based on Neural Networks. *Journal of Computer and Communications*, **14**, 1-9.
https://doi.org/10.4236/jcc.2026.149001

Received: August 6, 2026
Accepted: September 13, 2026
Published: September 16, 2026

Abstract

This article contributes to improving communication between the deaf community and the rest of the population through the use of digital technology and artificial intelligence techniques, notably Deep Learning, by developing a system capable of recognizing isolated signs in text of Niger Sign Language (NiSL) gestures (words) in real time using a webcam. For this purpose, a video dataset specific to Niger Sign Language was created. Landmarks of articulations were extracted using MediaPipe Holistic [1] and used to train a Long Short-Term Memory (LSTM) classification model to perform this recognition of isolated signs in text. The results obtained show a good performance of the model, with a recognition rate of 62.22% when only hand characteristics are used, which is comparable to results of previous works in the literature on more documented sign languages.

Keywords

Niger Sign Language, Automatic Translation, Recurrent Neural Networks, LSTM

1. Introduction

Communication is a fundamental process in human interaction, enabling the exchange of ideas, information and emotions [2]. For the deaf community, often deprived of hearing or speech, non-verbal communication, and more specifically sign language, becomes the main vector of interaction.

In Niger, disabled people represent 4.2% of the population, approximately 715,497 individuals, among whom there are nearly 24,000 deaf people [3]. The latter use Niger Sign Language (NiSL) to communicate. Sign language is a natural language in its own right, which is based on a visual-gestural modality involving the hands, facial expressions and body movements [4]. However, in an environment where the spoken language predominates, this community faces social isolation which hinders its integration.

Technological advances, particularly in artificial intelligence (AI), offer promising solutions to overcome this isolation. Systems based on digital gloves (data gloves) [5] or computer vision [6] have been developed for other sign languages. However, NiSL remains underrepresented in academic work and technological applications, suffering from a critical lack of digital resources.

This work aims to address this challenge by developing a system capable of recognizing NiSL gestures at word level in real time and translating them into text, using a simple webcam.

2. Materials and Methods

To design this translation system, we adopted a multi-step methodology, ranging from the linguistic study of the NiSL to the validation of the AI model.

2.1. Linguistic Aspects of Sign Language

Contrary to popular belief, sign languages are not simple gestural transcriptions of spoken languages. They are complex languages with their own phonology, morphology, and syntax. The basic unit, analogous to the phoneme, is the sign, which is composed of several essential parameters. William Stokoe [4] was the first to identify three of such parameters for American Sign Language (ASL):

- Hand configuration: The shape the hand takes (e.g., closed fist, open hand).
- Movement: The movement of the hand in space (direction, speed).
- Location: The place where the sign is produced on the body or in space.

Other researchers have complemented this model by adding palm orientation and non-manual elements (facial expressions, body movements), which are crucial for nuanced meaning [7]. Our modeling approach aims to capture this multimodal richness.

2.2. Construction of a Data Set for NiSL

One of the major challenges of this project was the lack of a digital corpus for the NiSL. To address this, we built our own dataset:

1) Source: We based our work on the Dictionary of Signs in use in Niger, a paper document used in schools for the deaf in Niamey. The corpus was recorded by a single experienced deaf signer, a deaf person from the Niamey School for the Deaf/Hassane Banâ Bâ, a public institution in Niamey. This signer provided all nine original recordings. Using a deaf signer instead of directly reproducing the dictionary illustrations is a methodological choice: the source only has static draw-

ings, and only a skilled signer can convey the movement, rhythm, and articulation that a drawing can't capture.

2) Acquisition: We recorded videos of 9 distinct signs related to the family estate. Each video was captured in high definition with a Canon LEGRIA HF R806 camera to ensure accurate gesture analysis.

3) Data enrichment: To increase the robustness and diversity of our initial corpus of 9 videos, we applied data augmentation techniques:

- **Horizontal Flip** [8]: To simulate the variability between right-handed and left-handed signers, each video was duplicated in a mirror version, bringing the corpus to 18 videos.
- **Temporal Variation:** For each video, we generated 15 new instances with variable execution speeds (slowed down and accelerated).

In the end, the dataset consists of **270 videos** (9 signs × 2 hand variations × 15 speed variation) (see Figure 1).

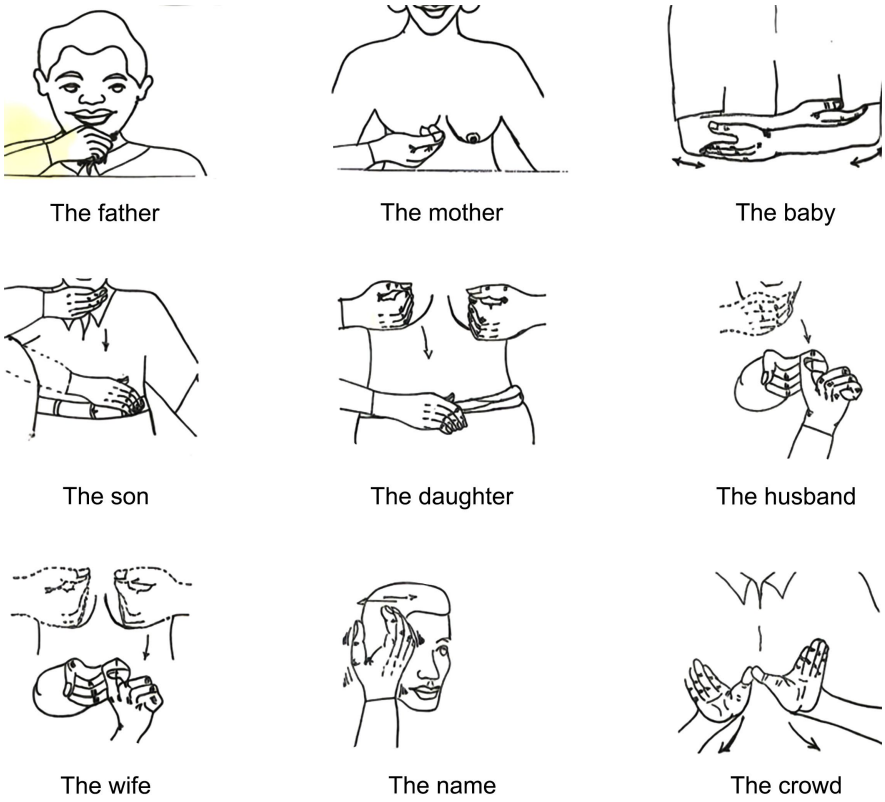

Figure 1. The 9 dictionary signs used.

3. Modeling and Design

The heart of this system is a Deep Learning model capable of processing sequential and multimodal data.

3.1. Feature Extraction with MediaPipe Holistic

To transform the raw video data into a format usable by a neural network, we used

Google's MediaPipe Holistic solution [1]. This powerful tool allows real time and simultaneous extraction of key points (landmarks) of the body, face and hands from each frame of the video.

- Hands: 21 landmarks per hand (see **Figure 2**)
- Face: 468 landmarks (see **Figure 3**)
- Pose: 33 landmarks (see **Figure 4**)

For each video frame, the feature vectors for the right hand $v(M_d)$, the left hand $v(M_g)$, the face $v(F)$ and the pose $v(P)$ are extracted. These vectors are then concatenated to form a single feature vector $v(I)$ for frame.

$$v(I) = v(M_d) \oplus v(M_g) \oplus v(F) \oplus v(P) \tag{1}$$

where $\oplus$ represents the concatenation operation. A video is thus represented by a sequence of these vectors V_{sign}

$$V_{\text{sign}} = \left(v(I_1), v(I_2), \cdots, v(I_N)\right) \tag{2}$$

where N denotes the total number of frames in the video.

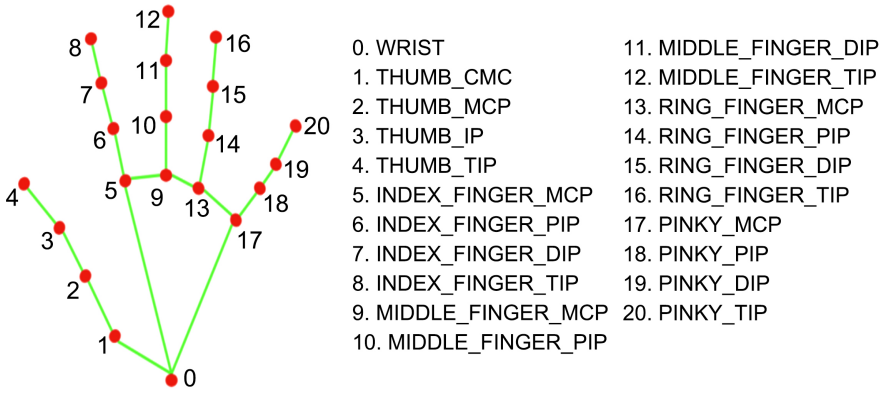

Figure 2. Landmarks of a hand [9].

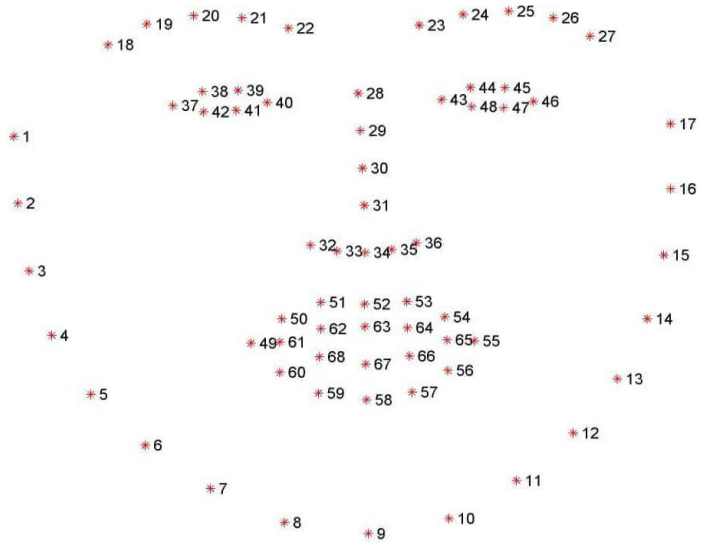

Figure 3. Landmarks of a face [10].

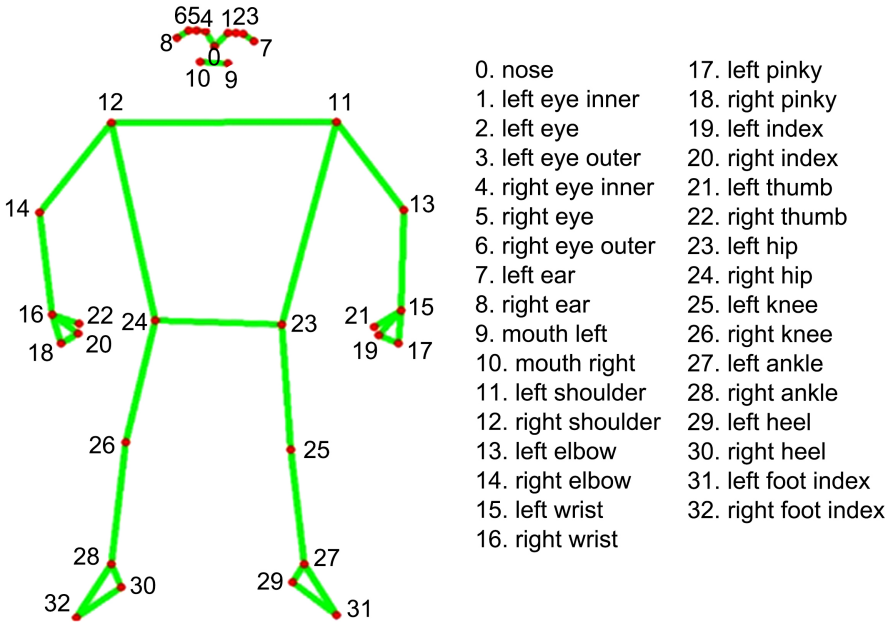

Figure 4. Landmarks of a pose [9].

3.2. LSTM Model Architecture

For the task of classifying gesture sequences, we opted for a recurrent neural network, and more specifically an LSTM (Long Short-Term Memory) architecture [11]. LSTMs are particularly effective at learning long-term dependencies in sequential data, which is essential for understanding the dynamic movements of signs.

The architecture of our model is as follows: (see Figure 5)

1) LSTM Layers: Two overlapping LSTM layers (64 and 128 neurons) to process the sequence of landmark vectors and learn temporal patterns.

2) Dense Layers (MLP): Two fully connected layers (64 neurons each) to interpret the features learned by the LSTMs.

3) Dropout Layer: A regularization layer to prevent overfitting.

4) Output Layer: A dense layer with a softmax activation function that produces a probability distribution over the nine (9) classes (the classes of the nine signs).

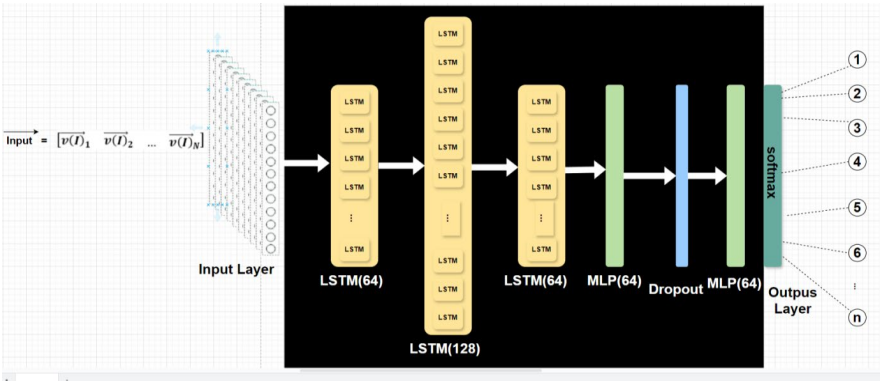

Figure 5. LSTM neural network architecture.

4. Results and Discussion

For the evaluation, we didn't use part of these 270 sequences. We did a new, independent real-time recording: each sign was performed 10 times, making 90 test trials, with no overlap with the training data. So, the reported performance is calculated exclusively on these 90 independent trials. We trained and tested the model in four different configurations to assess the relative importance of each modality (hands, face, pose). The model was tested in real time via a webcam, performing 10 attempts for each of the 9 signs.

The average recognition rates obtained are as follows: (see Table 1, Figure 6 and Figure 7)

- Hands only: 62.22%
- Hands + Face: 27.77%
- Hands + Pose: 35.55%
- Hands + Face + Pose: 37.77%

Table 1. Overall recognition rate based on features used.

Criteria	Hands only	Hand + Pose	Hand + Face	Hand + Face + Pose
Number of signs tested	9	9	9	9
Repetitions by sign	10	10	10	10
Correct predictions	56	32	25	34
Incorrect predictions	34	58	65	56
Total number of predictions	90	90	90	90
Accuracy rate	**62.22%**	**35.55%**	**27.77%**	**37.77%**

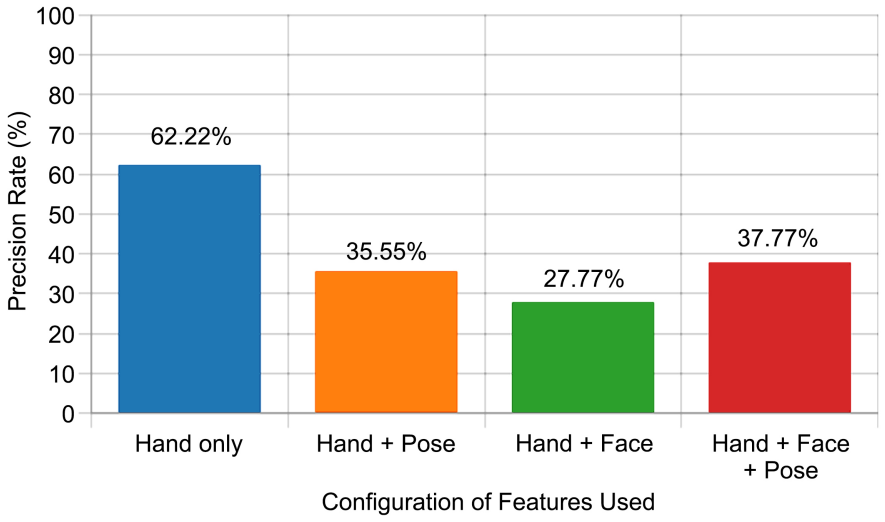

Figure 6. Comparison of the overall accuracy rates of the four configurations.

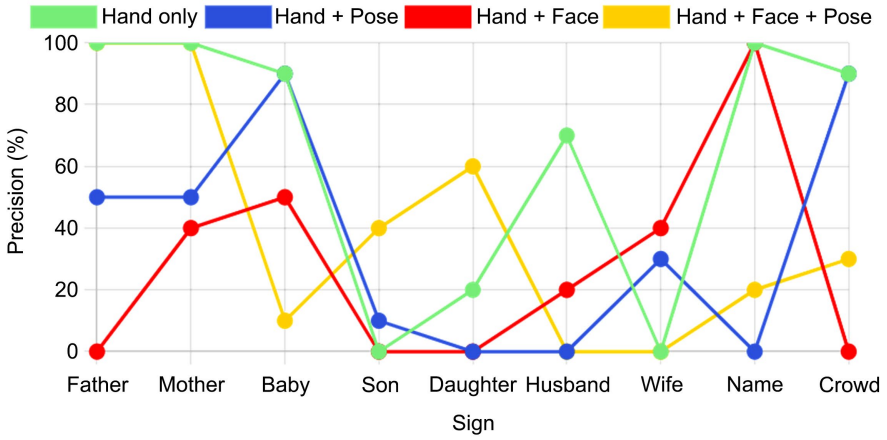

Figure 7. Accuracy by sign according to feature configuration.

These results suggest that using only hand features achieves the best recognition rate. This score reflects the central role of manual gestures, which convey the majority of meaning in sign language. The lower complexity of this configuration (fewer key points to process) allows the model to efficiently focus on the most relevant information.

On the other hand, adding facial features significantly reduces accuracy. This can be explained by the high variability of facial expressions and the large number of extracted key points (468), which may introduce noise and distract the model from more discriminative hand features.

Including pose features slightly improves accuracy compared to using facial features alone, likely because body posture provides useful additional spatial context. However, its contribution remains secondary compared to hand gestures.

Finally, combining all features only marginally improves performance over face and pose alone, while remaining well below the hand-only configuration. The increased data dimensionality and the absence of a mechanism to weight the relative importance of each modality appear to dilute key signals, reducing the overall effectiveness of the model.

Our best result (62.22%) is encouraging and comparable to those reported in the literature for more extensively studied sign languages. For instance, studies conducted on the WLASL dataset for American Sign Language (ASL) reported accuracies of 62.63% [12], although it should be noted that experimental conditions differ significantly in terms of vocabulary size and dataset scale.

It should also be noted that, due to the binary nature of the collected data (correct/incorrect prediction), only accuracy and recall could be computed in this study. Metrics such as precision, F1-score, and confusion matrix require per-prediction class labels, which constitute a limitation of the current evaluation and a direction for future work.

5. Conclusions

This work has enabled the development of a functional prototype of a real-time

translation system from Niger Sign Language to text, using Deep Learning techniques. The main contribution lies in the creation of a first video corpus for NiSL and in the demonstration that an LSTM model, trained on hand movements alone, can achieve a promising performance of 62.22%.

Our results emphasize that hands convey the essence of the message and that the addition of other modalities can, paradoxically, degrade performance due to the complexity and noise added.

To improve this system, future perspectives include the expansion of the corpus. The integration of attention mechanisms, such as those proposed in Transformers models [13], could also help to better prioritize information from different modalities. Ultimately, such a system could greatly promote the social and professional inclusion of deaf people in Niger.

Author Contributions

Conceptualization, Moussa Idi Bachir and Bouwey Alley Ibrahim; methodology, Moussa Idi Bachir; software, Bouwey Alley Ibrahim; validation, Naroua Harouna, Kadri Chaibou, and Moussa Idi Bachir; resources, Bouwey Alley Ibrahim; data curation, Bouwey Alley Ibrahim; writing—original draft preparation, Moussa Idi Bachir and Kadri Chaibou; writing—review and editing, Kadri Chaibou and Morou Ganda Yahaya; supervision, Naroua Harouna.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Grishchenko, I. and Bazarevsky, V. (2020) MediaPipe Holistic—Simultaneous Face, Hand and Pose Prediction, on Device. Google Research Blog.

[2] Littlejohn, S.W. and Foss, K.A. (2010) Theories of Human Communication. 10th Edition, Waveland Press.

[3] Studio Kalangou (2023) La langue des sourds: Un apport à l'éducation des personnes déficientes auditives et malentendantes au Niger.

[4] Stokoe, W.C. (1960) Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf. Studies in Linguistics: Occasional Papers 8, University of Buffalo.

[5] Ahmed, M.A., Zaidan, B.B., Zaidan, A.A., Salih, M.M. and Lakulu, M.M.b. (2018) A Review on Systems-Based Sensory Gloves for Sign Language Recognition State of the Art between 2007 and 2017. *Sensors*, **18**, Article 2208. https://doi.org/10.3390/s18072208

[6] Mitra, S. and Acharya, T. (2007) Gesture Recognition: A Survey. *IEEE Transactions on Systems, Man and Cybernetics, Part C* (*Applications and Reviews*), **37**, 311-324. https://doi.org/10.1109/tsmcc.2007.893280

[7] Crasborn, O.A. (2006) Nonmanual Structures in Sign Language. In: Brown, K., Ed., *Encyclopedia of Language & Linguistics*, Elsevier, 668-672. https://doi.org/10.1016/b0-08-044854-2/04216-4

[8] Shorten, C. and Khoshgoftaar, T.M. (2019) A Survey on Image Data Augmentation

for Deep Learning. *Journal of Big Data*, **6**, Article No. 60. https://doi.org/10.1186/s40537-019-0197-0

[9] GeeksforGeeks (2023) Python-Facial and Hand Recognition Using MediaPipe Holistic. GeeksforGeeks.

[10] Albadawi, Y., AlRedhaei, A. and Takruri, M. (2023) Real-Time Machine Learning-Based Driver Drowsiness Detection Using Visual Features. *Journal of Imaging*, **9**, Article 91. https://doi.org/10.3390/jimaging9050091

[11] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[12] Li, D.X., *et al.* (2020) WLASL: A Large-Scale Word-Level American Sign Language Dataset for Deep Learning-Based Recognition. https://dxli94.github.io/WLASL/

[13] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I. (2017) Attention Is All You Need. 31*st Conference on Neural Information Processing Systems* (*NIPS* 2017), Long Beach, CA, 1-11. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf