Classification of Dynamic Libras Signs Related to the Hospital Context Using Recurrent Neural Networks

Pedro de Araújo Bhering Bittencourt, Luiz Maurílio da Silva Maciel(✉), Marcelo Bernardes Vieira, and Saulo Moraes Villela

Department of Computer Science, Federal University of Juiz de Fora, Juiz de Fora, Brazil
pedro.bittencourt@estudante.ufjf.br, {luiz.maciel,marcelo.bernardes,saulo.moraes}@ufjf.br

Abstract. Communication between Brazilian healthcare professionals and Deaf individuals remains a significant challenge in hospital environments, especially due to the scarcity of professionals fluent in Brazilian Sign Language (Libras). In this context, sign recognition systems emerge as a promising alternative to reduce communication barriers. This work proposes the construction of a dataset composed of ten dynamic Libras signs related to hospital triage, as well as the development of recurrent neural network–based models for their classification. The signs are represented by three-dimensional landmarks extracted using MediaPipe Holistic, capturing spatial and temporal information of the movements, which are classified by LSTM networks. Initially, we investigated the 10-class classification problem, achieving an average accuracy of 83.83% and revealing frequent confusions between pairs of semantically similar signs. To address this limitation, we proposed a hierarchical strategy based on expert models, consisting of a general 7-class classifier combined with dedicated models for distinguishing frequently confused signs. This approach resulted in a significant improvement in overall performance, increasing the average accuracy to 92.44% and substantially reducing confusion errors between classes. Additional experiments conducted on an external subject, under less controlled acquisition conditions, achieved an average accuracy of 98.17%, further supporting the generalization capability of the proposed approach. These results indicate the potential of the method for communication support applications in healthcare environments.

Keywords: Machine learning · Social inclusion · Computer vision · Sign language

1 Introduction

Brazilian Sign Language (Libras) plays a fundamental role in promoting the social inclusion of the Deaf community by reducing communication barriers

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026, LNCS 16767, pp. 529–543, 2027.
https://doi.org/10.1007/978-3-032-30491-9_33

and ensuring the full exercise of citizenship. In Brazil, this right is guaranteed by the Brazilian Law for the Inclusion of Persons with Disabilities (Law No. 13,146/2015), which establishes equal access to essential services, including healthcare, through appropriate communication resources [1].

However, in the hospital context, communication between healthcare professionals and Deaf patients remains a significant challenge. The shortage of professionals trained in Libras compromises the quality of care and perpetuates inequalities in access to health services. Studies indicate that most medical students and healthcare professionals are not fluent in Libras, despite recognizing its relevance in clinical practice [2]. This scenario reveals a structural gap that directly affects the safety, autonomy, and dignity of Deaf patients.

Given these limitations, assistive technologies based on automatic sign recognition have shown promise in supporting communication in clinical environments. Several studies have explored the use of machine learning and neural networks for the classification of Libras signs, often accompanied by the construction of custom datasets [3–5]. However, most of these datasets are collected in controlled environments, with limited variability in participants and scenarios, which restricts the generalization capability of the models in real-world contexts, such as hospitals.

Recent approaches have sought to mitigate these limitations by representing gestures through body landmarks, which capture relevant spatial and temporal information in a compact and robust manner. This representation directly integrates with sequential models, such as Long Short-Term Memory (LSTM) neural networks, which have been widely employed in the interpretation of dynamic sign language gestures [6–8].

In this context, this work proposes the creation of a dataset composed of dynamic Libras signs related to the healthcare domain, with a focus on hospital triage, as well as the development and evaluation of recurrent neural network–based models for the classification of these signs. The proposed approach aims to contribute to the reduction of communication barriers in clinical environments and to the promotion of more inclusive healthcare for the Deaf community.

2 Related Work

In recent years, advances in deep learning have significantly increased research in sign language recognition, particularly in video-based approaches. Among the most promising strategies are the extraction of body landmarks – such as hands, face, and pose – and the use of temporal modeling architectures, including LSTM networks.

Huang and Chouvatut [8] proposed a video-based pipeline that combines a ResNet network for visual feature extraction with an LSTM layer to capture temporal dependencies, achieving competitive results when evaluated on the LSA64 dataset [9].

Gandhi, Shah, and Chandane [7] integrated the MediaPipe library for the automatic extraction of hand, face, and pose landmarks, using these vectors as input to an LSTM network for American Sign Language (ASL) recognition.

Samaan et al. [6] compared different recurrent architectures, including GRU, LSTM, and Bi-LSTM, demonstrating that LSTM-based architectures achieve superior performance in scenarios with higher temporal complexity.

In the Brazilian context, Sarmento [10] highlighted the effectiveness of using only landmarks in combination with LSTM networks for Libras sign classification, outperforming image-based approaches.

Other studies have also reported promising results by integrating spatial feature extraction techniques with temporal modeling [11,12], reinforcing the robustness of landmark-based representations and recurrent neural networks for dynamic sign language recognition.

In summary, the literature indicates that the combination of landmark-based representations and LSTM models constitutes an effective approach for dynamic sign classification. In this sense, the present work adopts a similar strategy, focusing on the recognition of Libras signs in the healthcare context.

3 Proposed Dataset

We did not identify public datasets that include signs from Brazilian Sign Language (Libras) specifically targeted at the hospital context. Thus, this work proposes the construction of a dedicated dataset to represent a more realistic application scenario in clinical environments.

The dataset is composed of 10 dynamic Libras signs related to hospital triage, performed by 9 participants fluent in Libras, with diversity in physical and visual characteristics. In addition, one participant was recorded on two occasions under different conditions – varying clothing, accessories (such as glasses), and hairstyle (tied or loose hair) – which contributed to increasing the diversity of the collected material. Each participant performed each sign up to five times, totaling 492 videos. We carried out the recordings in an indoor environment with controlled lighting, a fixed camera, and a uniform background in order to reduce external interference and ensure full visibility of hand and arm movements.

The selected signs – Allergy, Autism, Bronchitis, Cramp, Seizure, Diarrhea, Fracture, Athlete's Foot, Heart Attack, and Measles – are dynamic in nature and involve coordinated movements of the hands, arms, and body posture, thus requiring appropriate temporal modeling. These signs were chosen to represent common symptoms and medical conditions encountered in hospital settings. Videos were recorded at a spatial resolution of 1280×720 pixels (HD) or 640×480 pixels, with a temporal resolution of 30 frames per second (FPS).

This dataset provides an appropriate basis for evaluating Libras sign recognition models in the healthcare context, serving as experimental support for the methods proposed in this work.

4 Proposed Method

Initially, we applied preprocessing to the videos to standardize the model input. As the duration of sign execution varies between participants, we resampled each

video to contain 30 frames. Frame selection was performed using a sampling strategy based on a normal distribution centered at the midpoint of the video duration. This strategy prioritizes frames with the most relevant movements without completely disregarding the remaining parts of the sequence.

To reduce overfitting and increase the variability of the dataset, we applied data augmentation. The geometric transformations used included mild rotations, horizontal and vertical translations, and spatial cropping, which were consistently applied to all frames of each video to preserve the temporal coherence of the gesture. Operations such as horizontal flipping and color transformations were not used to avoid semantic distortions of the signs and potential interference with the landmark extraction process.

Furthermore, these augmentation operations were integrated into the frame sampling process. This procedure was repeated 20 times for each video. As a result, each video produced multiple variations of the same gesture, contributing to more diverse and representative cases.

After preprocessing, body and hand landmarks were extracted using MediaPipe Holistic, which enables the simultaneous detection of these regions with spatial and temporal consistency. Although facial landmarks were available, they were discarded because facial expressions were not sufficiently perceptible in some of the videos, which could introduce inconsistencies into the data.

For each frame, 75 three-dimensional landmarks were considered: 33 correspond to body pose and 42 to the hands (21 for each hand). We concatenated these points into a single one-dimensional vector containing 225 values per frame. Thus, each video was represented by a tensor of size 30×225, which we used as input to the recurrent neural network. To optimize processing and avoid repeating the extraction step, all feature vectors were stored in binary files using the NumPy library.

Figure 1 illustrates an example of pose and hand landmarks estimated on a video frame from the dataset used in this work. Although the visualization is presented in two dimensions, the extracted points correspond to three-dimensional coordinates and preserve the temporal structure of the gesture throughout the sequence.

The model proposed in this work is composed of four main layers: Input, Masking, LSTM, and Classification. This architecture was designed to handle sequential data obtained from landmark extraction while preserving the temporal dependencies of the performed gestures.

Initially, the flattened landmark vectors are fed into the Input layer. Next, a Masking layer is applied to ignore null values resulting from occasional failures in the extraction process, preventing such information from influencing the model's learning.

The core of the architecture consists of an LSTM recurrent network, which is responsible for capturing the temporal dependencies present in the frame sequences. Other recurrent architectures, including Gated Recurrent Units (GRU) and bidirectional LSTM variants, were also evaluated during model development. However, these alternatives did not show consistent improvements

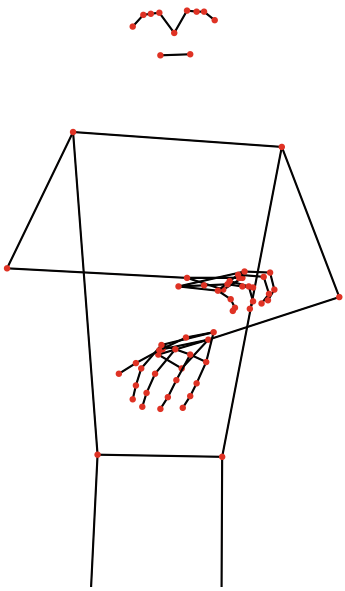

Fig. 1. Pose and hand landmarks estimated by MediaPipe Holistic.

over the standard LSTM in terms of average accuracy or validation stability. Therefore, LSTM was chosen as the primary architecture, offering the best balance between performance, complexity, and generalization. This layer employs the ReLU activation function and uses recurrent dropout as a regularization technique. The application of dropout directly to recurrent connections follows well-established approaches in the literature for recurrent neural networks, contributing to the reduction of overfitting without compromising the modeling of long-term dependencies [13].

Finally, a fully connected layer performs the classification, producing the probability that the video belongs to each of the ten sign classes considered in this work.

5 Experiments and Results

We conducted the experiments on a machine equipped with an Intel(R) Core(TM) i5-10300H CPU at 2.50 GHz, 16 GB of RAM, and running Windows 11, with the implementations executed within the Windows Subsystem for Linux (WSL2) using Ubuntu 22.04.5 LTS. All models were implemented in Python (version 3.10.2) using the TensorFlow library.

The dataset was evaluated using the k-fold cross-validation technique with $k = 10$, following the leave-one-subject-out scheme. In this procedure, at each iteration, the data of one individual are used exclusively for testing, while the data of the remaining individuals compose the training set. Taking into account

the different recording conditions adopted, we treated each occurrence of a subject recorded twice as a distinct subject, resulting in a total of ten subjects in the dataset.

In all experiments, the AdamW optimizer was employed, configured with a weight decay of 0.005. The hyperparameters related to the learning rate, the number of neurons in the LSTM layer, and the recurrent dropout rate vary among the evaluated models.

In addition, each cross-validation fold was evaluated 20 times. This strategy arises from the frame selection process based on sampling from a normal distribution, which can generate a distinct version of each video for each sample. Thus, for each subject used as the test set, we obtained 20 independent evaluations, leading to more consistent results.

Note that these 20 samples correspond to the versions generated during the frame sampling process. The sampled sequences corresponding to the test subject were used without any modification. However, for the training subjects, these sequences were used to generate augmented variations. The performance of the models was assessed using the metrics: accuracy, precision, recall, F1 score, and F2 score. We evaluated each sample and reported the mean values.

5.1 10-Class Classification Problem

For the classification problem encompassing the 10 classes of the dataset, the model that achieved the best performance was configured with a batch size of 128, a learning rate of 0.001, 64 neurons in the LSTM layer, a recurrent dropout rate of 0.2, and 300 training epochs. This model achieved an average general accuracy of 83.83%, with a standard deviation of 0.72%, indicating consistent performance across the different samplings.

Table 1 presents the mean values and standard deviations of the performance metrics per class achieved from the evaluation of the 20 samples used in the cross-validation procedure.

We observe that classes such as Autism, Bronchitis, Seizure, Diarrhea, and Measles achieved high performance with a relatively balanced relationship between precision and recall across the 20 test sets, indicating the good generalization capability of the model for these signs. The Diarrhea class stands out, maintaining mean recall values close to 98% and F1 scores above 98%, which evidences high gestural separability.

On the other hand, the Allergy, Cramp, and Athlete's Foot classes presented the lowest average performance, particularly in terms of recall, indicating a higher incidence of false negatives across the evaluated sets. In particular, the Cramp class exhibited a mean recall below 60%, while Athlete's Foot showed difficulties in both precision and recall, resulting in reduced F1 score values. These results highlight the structural limitations of the model in distinguishing these classes.

The average of all classes, around 84%, reinforces the presence of relevant performance asymmetries between classes. This indicates that despite satisfac-

Table 1. Performance evaluation per class for the 10-class classification problem.

Class	Precision	Recall	F1 score	F2 score
Allergy	73.86 ± 2.05	73.33 ± 3.30	73.53 ± 2.13	75.70 ± 2.77
Autism	94.89 ± 2.11	93.50 ± 2.44	94.15 ± 1.43	93.75 ± 1.92
Bronchitis	85.84 ± 1.32	98.96 ± 1.04	91.92 ± 0.94	96.02 ± 0.90
Cramp	85.77 ± 4.20	56.87 ± 3.30	68.31 ± 2.78	60.94 ± 3.07
Seizure	83.08 ± 3.27	90.10 ± 0.77	86.42 ± 1.89	88.58 ± 1.07
Diarrhea	99.80 ± 0.61	97.86 ± 1.76	98.81 ± 0.89	98.23 ± 1.40
Fracture	71.61 ± 2.15	95.10 ± 2.05	81.68 ± 1.77	89.23 ± 1.75
Athlete's Foot	64.36 ± 2.62	55.90 ± 3.32	59.79 ± 2.61	57.38 ± 3.01
Heart Attack	98.95 ± 1.16	83.50 ± 2.09	90.55 ± 1.32	86.18 ± 1.79
Measles	85.74 ± 2.50	93.62 ± 1.65	89.49 ± 1.82	91.92 ± 1.61
Average	84.39 ± 0.71	83.87 ± 0.72	83.47 ± 0.76	83.56 ± 0.75

tory overall performance, certain categories negatively impact the quality of the model.

Figure 2 shows the confusion matrix obtained for the test set that achieved the highest overall accuracy, namely 84.96%. We note recurrent error patterns between specific pairs of signs, such as Allergy and Athlete's Foot, Bronchitis and Heart Attack, and Cramp and Fracture. These confusions are attributed to intrinsic gestural similarities between the signs, which motivated the reformulation of the problem through class merging and the development of expert models to perform refined classification within each group. This approach leads to the development of a 7-class classification model and three expert models, each designed to handle a specific pair of confused classes.

5.2 7-Class Classification Problem

For the 7-class classification problem, the best-performing experiment used a batch size of 128, a learning rate of 0.001, 32 neurons in the LSTM layer, a recurrent dropout rate of 0.2, and 400 training epochs. To ensure class balance after unification, 10 samples from each original class were selected prior to merging, while the test sets were kept unchanged.

Table 2 presents the mean values and standard deviations of the performance metrics obtained from the evaluation of the 20 test sets used in the cross-validation procedure for the 7-class problem.

Note that all classes achieved mean values greater than 90% for precision, recall, F1, and F2 scores, indicating high and stable performance. The average of all classes was close to 96%, showing balanced sensitivity and precision between classes.

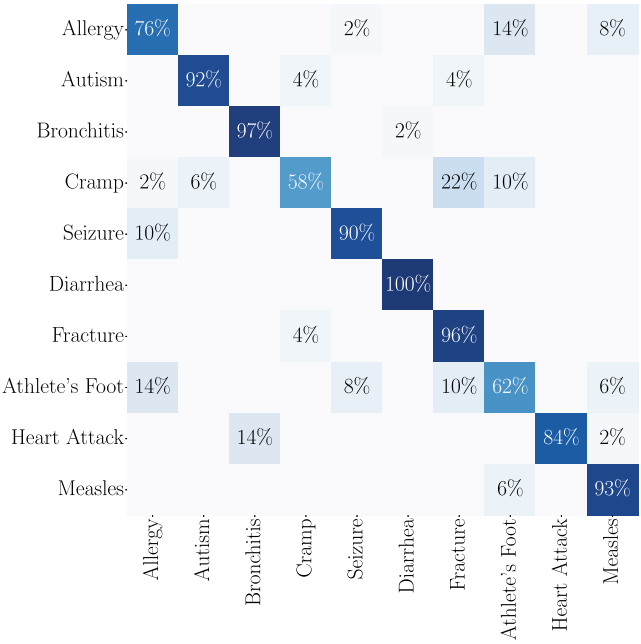

Fig. 2. Confusion matrix for the original 10-class classification problem.

5.3 Classification Problems Using Expert Models

Autism, Bronchitis+Heart Attack, and Cramp+Fracture classes performed particularly well, with mean metrics near 100%, reflecting the improved separability after merging previously confusing classes.

After the 7-class model, we developed the model responsible for classifying the signal between the Bronchitis and Heart Attack classes, as well as the model responsible for classifying the signal between the Cramp and Fracture classes. This choice was motivated by the fact that both followed the same development methodology described in the previous processes.

For the model that classifies signals between Bronchitis and Heart Attack, the experiment that achieved the best performance used a batch size of 32, a learning rate of 0.001, 32 neurons in the LSTM layer, a recurrent dropout rate of 0.2, and 150 training epochs. This configuration achieved a mean accuracy of 93.42% across the evaluated test sets, with a standard deviation of 0.60.

Table 3 presents the mean values and standard deviations of the performance metrics achieved from the evaluation of the 20 test sets for the Bronchitis and Heart Attack signal classification problem.

By comparing the results in Table 3 with the 10-class scenario (Table 1), we can observe a clear improvement in the balance between precision and recall. In the original model, the Heart Attack class had reduced recall due to frequent confusion with Bronchitis, an issue that was mitigated in the expert model.

Table 2. Performance evaluation per class for the 7-class classification problem.

Class	Precision	Recall	F1 score	F2 score
Allergy+Athlete's Foot	91.93 ± 1.02	91.00 ± 1.34	91.46 ± 0.86	91.18 ± 1.10
Autism	99.71 ± 0.70	100.00 ± 0.00	99.85 ± 0.35	99.94 ± 0.14
Bronchitis+Heart Attack	99.29 ± 0.72	99.08 ± 0.64	99.18 ± 0.57	99.12 ± 0.58
Cramp+Fracture	97.82 ± 0.95	98.11 ± 1.03	97.96 ± 0.70	98.05 ± 0.85
Seizure	97.28 ± 1.78	88.40 ± 1.36	92.62 ± 1.16	90.04 ± 1.20
Diarrhea	98.15 ± 1.26	96.73 ± 1.76	97.43 ± 1.12	97.01 ± 1.45
Measles	88.23 ± 1.42	99.47 ± 1.14	93.50 ± 1.06	96.99 ± 1.02
Average	96.06 ± 0.50	96.11 ± 0.45	96.00 ± 0.46	96.05 ± 0.45

Table 3. Performance evaluation per class for the Bronchitis and Heart Attack signal classification problem.

Class	Precision	Recall	F1 score	F2 score
Bronchitis	89.54 ± 0.85	98.02 ± 0.80	93.59 ± 0.58	96.20 ± 0.64
Heart Attack	97.92 ± 0.84	89.00 ± 1.00	93.24 ± 0.63	90.65 ± 0.83
Average	93.73 ± 0.59	93.51 ± 0.60	93.41 ± 0.60	93.42 ± 0.60

As a result, Bronchitis maintained a high recall while showing increased precision, and Heart Attack achieved more balanced performance metrics, reflected in the F1 and F2 scores, indicating a reduction in confusion errors between classes.

For the model that classifies signals between Cramp and Fracture, the best-performing experiment used a batch size of 64, a learning rate of 1×10^{-4}, 32 neurons in the LSTM layer, a recurrent dropout rate of 0.2, and 150 training epochs. This configuration achieved an average accuracy of 90.87% across the evaluation sets, with a standard deviation of 2.18.

Table 4 presents the means and standard deviations of the performance metrics achieved from the evaluation of the 20 test sets for the Cramp and Fracture classification problem.

Table 4. Performance evaluation per class for the Cramp and Fracture signal classification problem.

Class	Precision	Recall	F1 score	F2 score
Cramp	94.65 ± 1.59	86.25 ± 4.29	90.19 ± 2.56	87.77 ± 3.61
Fracture	87.96 ± 3.28	95.30 ± 1.45	91.44 ± 1.88	93.71 ± 1.33
Average	91.30 ± 1.92	90.78 ± 2.21	90.82 ± 2.22	90.74 ± 2.25

Table 4 indicates that the expert model for the Cramp and Fracture signals achieved high and well-balanced performance, with consistent values of precision, recall, F1 score, and F2 score for both classes.

Compared to the 10-class scenario (Table 1), we note a significant improvement, particularly for the Cramp class, which previously exhibited low recall. The expert model substantially mitigated this limitation, while the Fracture class maintained high recall and achieved gains in precision, resulting in greater classification stability.

After resolving the first two pairs of unified classes, we developed an expert model to classify between the Allergy and Athlete's Foot signals. Unlike the previous cases, we observed that the main distinction between these signals occurs at the beginning of the movement. Therefore, we proposed a strategy for the automatic detection of gesture onset based on the temporal analysis of hand landmarks.

This approach consists of estimating motion energy from the temporal variation in the vertical positions of the hand centers. For each valid frame, the center of each hand is computed as the mean of its landmark coordinates. The resulting vertical trajectories are then smoothed using a Savitzky–Golay filter [14] to reduce frame-to-frame noise. Hand velocities are obtained via temporal gradients, and motion energy is defined as the sum of the magnitudes of the vertical velocities of both hands.

To detect the start of the gesture, we employ an adaptive threshold based on the initial behavior of the signal. The first five frames are assumed to represent a resting phase, and the threshold is defined as the mean energy within this window plus one standard deviation. The onset is identified as the first frame at which the energy remains above this threshold for five consecutive frames, thereby preventing isolated fluctuations from being misclassified as motion.

If the detected onset occurs within this initial window, a second threshold is estimated using frames 10 to 20. In this case, a more restrictive criterion is adopted, defined as the mean plus two standard deviations of the energy. The detection process is then repeated, ensuring greater robustness in sequences with noisy or unstable initial behavior.

Once the gesture onset is determined, the subsequent 30 frames are selected to compose the input for classification. These frames are uniformly sampled according to the video frame rate, ensuring temporal consistency of the gesture representation.

The best-performing model used a batch size of 32, a learning rate of 0.001, 128 neurons in the LSTM layer, a recurrent dropout rate of 0.1, and 250 training epochs. Cross-validation resulted in an overall accuracy of 97.00%, with 98.00% for the Allergy class and 96.00% for Athlete's Foot, indicating high and consistent performance.

Table 5 presents the precision, recall, F1 score, and F2 score metrics achieved for this classification problem. We observe a balanced behavior between the classes, with values above 96% across all metrics, indicating that the proposed strategy was effective in reducing the confusions observed in the 10-class model.

Table 5. Performance evaluation per class for the Allergy and Athlete's Foot signal classification problem.

Class	Precision	Recall	F1 score	F2 score
Allergy	96.08	98.00	97.03	97.61
Athlete's Foot	97.96	96.00	96.97	96.39
Average	97.02	97.00	97.00	97.00

5.4 Integration of 7-Class Model with Expert Models

Finally, the integration of the 7-class model with the expert models was performed. This combination resulted in an average accuracy of 92.44% in the 20 test sets, with a standard deviation of 0.78.

Table 6 presents the means and standard deviations of the performance metrics obtained after applying the expert-based strategy. This table enables a direct comparison with the results of the original 10-class model (Table 1).

Table 6. Performance evaluation per class using the expert model strategy.

Class	Precision	Recall	F1 score	F2 score
Allergy	82.45 ± 2.32	88.50 ± 1.77	85.35 ± 1.77	87.21 ± 1.67
Autism	99.71 ± 0.70	100.00 ± 0.00	99.85 ± 0.35	99.94 ± 0.14
Bronchitis	88.20 ± 1.40	96.35 ± 1.45	92.09 ± 1.14	94.60 ± 1.25
Cramp	89.85 ± 2.94	82.71 ± 5.15	86.05 ± 3.50	84.00 ± 4.46
Seizure	97.28 ± 1.78	88.40 ± 1.36	92.62 ± 1.16	90.04 ± 1.20
Diarrhea	98.15 ± 1.26	96.73 ± 1.76	97.43 ± 1.12	97.01 ± 1.45
Fracture	88.01 ± 3.23	95.10 ± 1.61	91.38 ± 1.96	93.57 ± 1.50
Athlete's Foot	97.79 ± 0.07	88.60 ± 2.69	92.95 ± 1.52	90.29 ± 2.25
Heart Attack	97.91 ± 0.84	88.80 ± 1.17	93.13 ± 0.72	90.48 ± 0.97
Measles	88.23 ± 1.42	99.47 ± 1.14	93.50 ± 1.06	96.99 ± 1.02
Average	92.76 ± 0.75	92.47 ± 0.79	92.43 ± 0.80	92.41 ± 0.80

Compared to the initial 10-class problem, in which the F1 score was around 83% (average of all classes), the strategy based on expert models increased this value to approximately 92%, accompanied by consistent improvements in average precision and recall. These results indicate a general enhancement in discriminative capacity.

The greatest gains are concentrated in the classes that exhibited greater gestural overlap in the original scenario, particularly the Allergy+Athlete's Foot and Cramp+Fracture pairs, whose performance metrics became more balanced after the application of expert models. In contrast, previously well-defined classes

such as Autism, Diarrhea, and Measles maintained good performance levels, demonstrating that the proposed approach did not compromise their recognition.

Figure 3 presents the confusion matrix for the same sample that generated the confusion matrix shown in Fig. 2. We can observe a significant reduction in class confusions, reflected in the increase in overall accuracy from 84.96% to 93.50%. Thus, decomposing the problem into expert classifiers has resulted in a more consistent system, outperforming the original model both in terms of average performance and in reducing ambiguities among semantically similar classes.

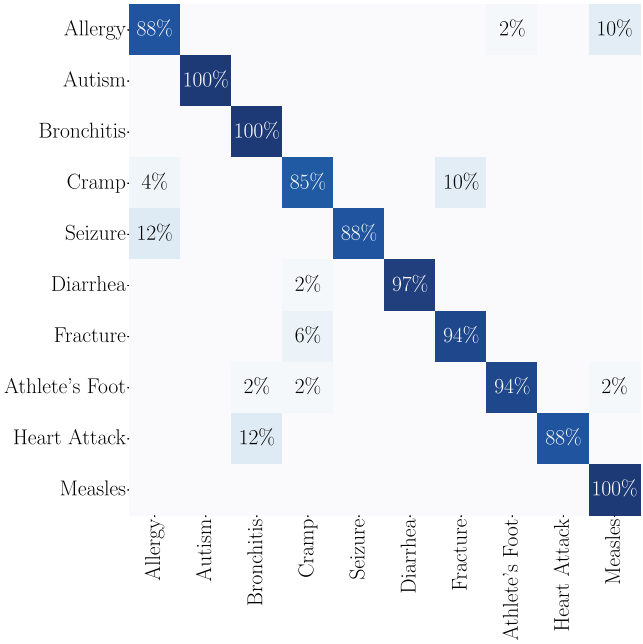

Fig. 3. Confusion matrix for the expert-model strategy.

5.5 Additional Test Set

In addition to the dataset collected with Libras professionals, an additional test set was recorded with a participant who is also fluent in Brazilian Sign Language.

We recorded a total of 60 videos, corresponding to 6 samples for each sign. Unlike the controlled acquisition process adopted for the main dataset, these recordings were performed under uncontrolled conditions, including variations in background and lighting. Furthermore, a different recording device was used, namely a Samsung S20 FE smartphone, with a resolution of 1080×1920 pixels and a frame rate of 60 FPS.

Despite these differences, the recording dynamics and the segmentation of individual sign executions followed the same procedures described for the main dataset.

This additional dataset is used to evaluate the generalization capability of the models trained on the original dataset, simulating more realistic usage conditions.

To ensure consistency with the evaluation protocol adopted in the previous experiments, we also performed 20 independent evaluations on this additional test set.

Table 7 presents the performance metrics obtained for this additional test set using the expert model strategy. The results show a general average accuracy of 98.11%, with a standard deviation of 1.66%, and all average metrics above 98%. This improvement can be attributed to the reduced variability of the dataset, which consists of a single participant, as well as the consistency in gesture execution.

Table 7. Performance evaluation per class using the expert model strategy in the additional test set.

Class	Precision	Recall	F1 score	F2 score
Allergy	98.57 ± 4.29	100.00 ± 0.00	99.23 ± 2.31	99.68 ± 0.97
Autism	99.29 ± 3.11	100.00 ± 0.00	99.62 ± 1.68	99.84 ± 0.70
Bronchitis	90.42 ± 9.34	99.17 ± 3.63	94.33 ± 5.75	97.10 ± 3.87
Cramp	96.43 ± 6.19	100.00 ± 0.00	98.08 ± 3.33	99.19 ± 1.40
Seizure	100.00 ± 0.00	96.67 ± 6.67	98.18 ± 3.64	97.24 ± 5.52
Diarrhea	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
Fracture	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
Athlete's Foot	100.00 ± 0.00	95.83 ± 7.22	97.73 ± 3.94	96.55 ± 5.97
Heart Attack	100.00 ± 0.00	90.00 ± 11.06	94.36 ± 6.43	91.63 ± 9.36
Measles	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
Average	98.47 ± 1.35	98.17 ± 1.66	98.15 ± 1.68	98.12 ± 1.71

Despite being recorded under uncontrolled conditions, the results indicate that the proposed approach is capable of generalizing well to new users and acquisition setups. However, it is important to note that the limited number of participants may simplify the classification task, and further evaluation with a larger and more diverse set of participants is necessary to comprehensively assess the generalization capability of the proposed method in real-world environments.

6 Conclusion

In this work, we proposed a method for the classification of Brazilian Sign Language (Libras) signs related to the healthcare domain. To this end, we constructed a dedicated dataset composed of 10 dynamic Libras signs associated

with hospital triage scenarios. Based on this dataset, we developed machine learning models using LSTM to handle temporal motion sequences and adequately represent the dynamic nature of the signs.

We observed that the model considering the 10 classes of the dataset presented recurrent confusions between specific pairs of signs, achieving an average accuracy of 83.83%. To address these confusions, we proposed a strategy based on expert models, restructuring the problem into a general 7-class classifier combined with dedicated classifiers for the most conflicting sign pairs. With this approach, the average accuracy increased to 92.44%, along with a substantial reduction in misclassifications between semantically similar signs. In addition, the low standard deviation observed across the test sets indicates that the system maintains consistent performance when exposed to new instances, reinforcing its practical potential for communication support applications in healthcare environments.

Additionally, an evaluation conducted on an additional set comprising an external subject not included in the original dataset, demonstrated that the proposed approach is capable of generalizing to new users and acquisition conditions. Despite variations in recording setup and environment, the model maintained high performance, reinforcing its robustness and potential applicability in real-world healthcare scenarios. Nevertheless, further validation with a larger and more diverse population remains necessary to fully assess its generalization capability.

In future works, it is possible to expand the sign set, include a greater diversity of users, and investigate more recent architectures, such as models based on attention mechanisms or vision transformers, with the aim of further improving performance and generalization capacity.

Acknowledgments. The authors would like to thank the Minas Gerais State Research Foundation - FAPEMIG, Brazil. This study was financed in part by the Federal University of Juiz de Fora - UFJF.

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

References

1. BRASIL: Lei nº 13.146, de 6 de julho de 2015. Diário Oficial da União: seção 1, Brasília, DF, 7 jul. 2015 (2015). https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13146.htm, institui a Lei Brasileira de Inclusão da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência). Accessed on: September 5, 2025. (in Portuguese)
2. Oliveira, A.S.R., Oliveira, C.V., Jesus, R.F., Quintanilha, L.F., Avena, K.M.: Ensino da língua brasileira de sinais durante a graduação em medicina: a percepção dos futuros médicos. Audiol. - Commun. Res. **27** (2022). https://doi.org/10.1590/2317-6431-2022-2634, fapUNIFESP (SciELO). (in Portuguese)

3. Rocha, J., Lensk, J., Ferreira, T., Ferreira, M.: Towards a tool to translate Brazilian sign language (libras) to Brazilian Portuguese and improve communication with deaf. In: 2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), pp. 1–4. IEEE (2020). https://doi.org/10.1109/VL/HCC50065.2020.9127257
4. Costa, C.F.F., Souza, R.S., Santos, J.R., Santos, B.L., Costa, M.G.F.: A fully automatic method for recognizing hand configurations of Brazilian sign language. Res. Biomed. Eng. **33**, 78–89 (2017). https://doi.org/10.1590/2446-4740.03816
5. Rezende, T.M., Almeida, S.G.M., Guimarães, F.G.: Development and validation of a Brazilian sign language database for human gesture recognition. Neural Comput. Appl. **33**(16), 10449–10467 (2021). https://doi.org/10.1007/s00521-021-05802-4
6. Samaan, G.H., et al.: Mediapipe's landmarks with RNN for dynamic sign language recognition. Electronics **11**(19), 3228 (2022). https://doi.org/10.3390/electronics11193228, https://www.mdpi.com/2079-9292/11/19/3228
7. Gandhi, D., Shah, K., Chandane, M.: Dynamic sign language recognition and emotion detection using mediapipe and deep learning. In: 13th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1–6. IEEE, Virtual Conference (2022). https://doi.org/10.1109/ICCCNT54827.2022.9984592
8. Huang, J., Chouvatut, V.: Video-based sign language recognition via resnet and LSTM network. J. Imaging **10**(6), 149 (2024). https://doi.org/10.3390/jimaging10060149, published: 20 June 2024
9. Ronchetti, F., Quiroga, F., Estrebou, C., Lanzarini, L., Rosete, A.: Lsa64: an Argentinian sign language dataset (2023). https://doi.org/10.48550/arXiv.2310.17429, instituto de Investigación en Informática LIDI, Facultad de Informática, Universidad Nacional de La Plata; Instituto Superior Politécnico José Antonio Echeverría
10. Sarmento, A.H.A.: Integração de Datasets de Vídeo para Tradução Automática da LIBRAS com Aprendizado Profundo. Dissertação de mestrado, Universidade de São Paulo, São Paulo, SP (2023). https://doi.org/10.11606/D.55.2023.tde-10012024-093541. (in Portuguese)
11. Ma, D., Hirota, K., Dai, Y., Jia, Z.: Dynamic sign language recognition based on improved residual-LSTM network. In: The 7th International Workshop on Advanced Computational Intelligence and Intelligent Informatics (IWACIII2021). Beijing, China (2021). https://iwaciii2021.bit.edu.cn/docs/2021-12/9caf45c507224bdca4933ad139fff639.pdf
12. Ridwang, Ilham, A.A., Nurtanio, I., Syafaruddin: Dynamic sign language recognition using mediapipe library and modified LSTM method. Int. J. Adv. Sci. Eng. Inf. Technol. (IJASEIT) **13**(6), 2171–2180 (2023). https://doi.org/10.18517/ijaseit.13.6.19401
13. Semeniuta, S., Severyn, A., Barth, E.: Recurrent dropout without memory loss (2016). https://doi.org/10.48550/arXiv.1603.05118
14. Konstantinovsky, T.: Introduction to the Savitzky–Golay filter: a comprehensive guide (using python) (2024). https://medium.com/pythoneers/introduction-to-the-savitzky-golay-filter-a-comprehensive-guide-using-python-b2dd07a8e2ce. Accessed 22 Oct 2025