Facial Expression Analysis in Brazilian Sign Language for Sign Recognition

Rúbia Reis Guerra[1], **Tamires Martins Rezende**[2], **Frederico Gadelha Guimarães**[2], **Sílvia Grasiella Moreira Almeida**[3]

[1]Departamento de Engenharia Elétrica – Universidade Federal de Minas Gerais Belo Horizonte – Minas Gerais – Brasil

[2]Programa de Pós-Graduação em Engenharia Elétrica – Universidade Federal de Minas Gerais – Belo Horizonte – Minas Gerais – Brasil

[3]Instituto Federal de Minas Gerais – Campus Ouro Preto – Ouro Preto – Minas Gerais – Brasil

{rubia-rg,rezendetamires,fredericoguimaraes}@ufmg.br
silvia.almeida@ifmg.edu.br

Abstract. *Sign language is one of the main forms of communication used by the deaf community. The language's smallest unit, a "sign", comprises a series of intricate manual and facial gestures. As opposed to speech recognition, sign language recognition (SLR) lags behind, presenting a multitude of open challenges because this language is visual-motor. This paper aims to explore two novel approaches in feature extraction of facial expressions in SLR, and to propose the use of Random Forest (RF) in Brazilian SLR as a scalable alternative to Support Vector Machines (SVM) and k-Nearest Neighbors (k-NN). Results show that RF's performance is at least comparable to SVM's and k-NN's, and validate non-manual parameter recognition as a consistent step towards SLR.*

1. Introduction

The first studies of sign language structure date back to 1960, with Stokoe [Landar and Stokoe 1961]. Sign language is a form of visual-motor communication used by the deaf community. Its smallest unit, a "signal", comprises non-manual parameters, movement of the face, eyes, head and torso, and manual parameters such as hand configuration, palm orientation, location and movement. Similar to spoken languages, sign languages have distinct grammatical structures, varying by country and culture [Gesser 2009]. According to [Laborit 1998], any concept can be expressed by means of signals without any loss of content.

Although the first mentions of sign language dissemination in Brazil date back to the 19th century [de Assis Silva 2012], just recently in Brazil it has been sanctioned as an official language. Brazilian Sign Language (LIBRAS) was only recognized in 2002 as the country's second official language, by law number 10.436 [Brasil 2002].

In the past decade alone, the improvement of machine learning techniques has led to significant advancements in automatic speech recognition. Speech-based Natural User Interfaces (NUI) were made possible and widely spread, facilitating human-human and human-machine interaction [López et al. 2017]. Despite recent progress, as seen

in [Hinton et al. 2012] and [Pigou et al. 2015], sign language recognition still lags behind. In particular, Brazilian sign language recognition has only recently been explored ([Rezende et al. 2016], [Freitas et al. 2017], [Filho et al. 2017]).

Most of the past work in LIBRAS recognition have focused primarily on the manual parameters of signals ([Almeida et al. 2014], [Dias et al. 2009], [Freitas et al. 2017]). This study is an extension of [Rezende et al. 2016], which proposes recognition of LIBRAS signs through facial expression. Facial components of signals are represented by the movement of the head, eye, eyebrow, etc., and are grammatical elements that make up the structure of the language, emphasizing and intensifying the signs when necessary.

There are over 10,000 signs in LIBRAS, and facial expressions are not mandatory in all signals [Capovilla 2017]. Moreover, different signals can share similar expressions. Hence, realizing signal classification tasks solely based on facial expression, in which each output label corresponds to a sign, does not generalize well for a large vocabulary. [Almeida et al. 2014] proposes a more scalable solution by extracting structural components of a signal, such as hand configuration and type of movement. Classification is then performed within each component, limiting the range of possible output classes. Signs can be recognized among others by grouping the individual results obtained for each component and comparing to a predetermined reference, such as [Capovilla 2017]. A similar approach is proposed in this work: instead of considering each signal's direct meaning as an output class, here we propose that each facial expression is labeled according to the closest fundamental expression (neutral, happy, sad, angry, fearful, surprised, disgusted) or according to the most prominent feature (e.g. tongue out or sucked cheeks).

The dataset used in this work is the same as presented in [Rezende et al. 2016], and comprises a descriptor of facial and manual spatial coordinates, and summarized frames of each signals' recordings [Rezende et al. 2016]. In this work, only coordinates pertaining to facial points were considered. In addition to experiments utilizing facial coordinates, classification is performed on two novel descriptors:

- Points selected by inspection of video frames;
- Facial points which suffer most displacement throughout frames.

Past studies in LIBRAS recognition have employed Support Vector Machines (SVM), k-Nearest Neighbors (k-NN) and artificial neural networks for classification tasks ([Porfirio et al. 2013], [Filho et al. 2017], [Almeida et al. 2014], [Rao et al. 2017]). In this work, due to the high dimensionallity of the available data, Random Forest [Breiman 2001] was proposed as a classification method in the learning stage. The Random Forest algorithm implicitly performs feature selection [Breiman 2001] and have consistently shown robust results throughout a plethora of applications, including facial expression recognition [Pu et al. 2015]. Furthermore, the algorithm yields good accuracy results in classification tasks when compared to other state-of-the-art methods [Zhang et al. 2017], is fast to train and can easily be parallelized [Genuer et al. 2017], posing as a scalable candidate for learning tasks on a larger collection of LIBRAS signs.

2. Related Work

Most literature on signal language recognition deal with the relative configuration of the hands [Escobedo-Cardenas and Camara-Chavez 2015],

[Almeida et al. 2014], [de Paula Neto et al. 2015], [Pariwat and Seresangtakul 2017], [Uddin and Chowdhury 2016]. Most of these studies carry out recognition of the alphabet in their respective languages. Few works focus just on the information of the non-manual expressions [Freitas et al. 2017], [Rezende et al. 2016], [Uddin 2015], seeking to emphasize the importance of facial expressions in sentences and even the recognition of signs. Although they achieve high rates of recognition of expressions, the common methodologies proposed are adequate only to the set of data used. This therefore limits replicability, because there is no way to universalize the dataset in the literature. In addition, recognizing a sign using only one of its parameters is unfeasible, since multiple signs may share the same configuration of a parameter, while others do not use it at all.

One of the greatest challenges of SLR is dealing with all parameters simultaneously, since each sign language has its own unique grammatical structure and some signs may incorporate only a few parameters. Performing recognition using more than one parameter is the work of [Rao et al. 2017] and [Yang and Lee 2011].

In [Rao et al. 2017], a real-time signaling system was implemented using the frontal camera of a cell phone. Twenty signs, among words and letters, were tested. The videos were recorded at a rate of 30fps and each signal was subjected to pre-processing, segmentation and feature extraction techniques. At the end of the process, the signals were labeled using an Artificial Neural Network approach. The developed application returned an audio from the performed signal. Despite the well-structured methodology, the sample size was small. Another weakness of the system is the number of neurons in the hidden layer. The value was chosen by trial and error, preventing recognition of new, unseen samples, since the parameters of the system would have to be recalculated.

In the work of [Yang and Lee 2013], the manual segments of signals were identified and then analyzed in regard to their configurations. Facial expression was investigated if there were any ambiguities related to the analysis of the hands. The database used has 24 signals, of which 17 are related to the alphabet and 5, to facial expressions, making up 98 sentences from the American Sign Language. Recognition is therefore accomplished in separate steps for each parameter, hand or face.

As seen, works presented in the literature do not have a common methodological mechanism that is capable of evaluating the real situations that a user of sign language encounters. Overcoming this barrier enables the creation of a system for instantaneous recognition of signals, facilitating communication with others who do not know the language. Thus, the intention of this paper is to propose a generalizable approach to SLR, by focusing on classifying each signal's basic parameters, instead of attempting to extract its direct meaning.

2.1. Facial Expression Recognition

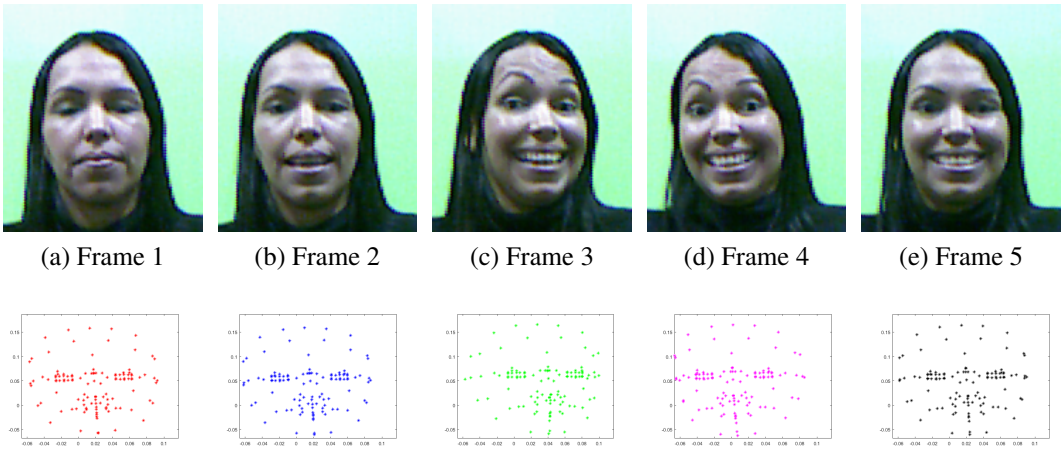
Automated analysis of human emotions is a multidisciplinary endeavor and a key component of human-computer interaction. A multitude of approaches have been studied in an effort to capture the nuances that differentiate facial expressions [Zeng et al. 2009].

The problem is well-studied and bench-marked within Computer Vision, presenting a variety of consolidated databases [Gross 2005]. Research directions differ, among a plethora of factors, in the choice of data, sentiment categorization, temporal or static

analysis and learning method ([Du et al. 2014], [Yu and Zhang 2015], [Jung et al. 2015], [Abdullah et al. 2014]).

2.2. LIBRAS Data

The data utilized in this work were obtained from [Rezende et al. 2016] and contain recordings of 10 different signs, represented by non-manual parameters of the Brazilian Sign Language (LIBRAS). Figure 1 shows five frames of a recording of the "happiness" sign. Each sign (to calm down, to accuse, to annihilate, to love, to gain weight, happiness, slim, lucky, surprise and angry) has 10 recordings, totaling a database of 100 samples.



(a) Frame 1 (b) Frame 2 (c) Frame 3 (d) Frame 4 (e) Frame 5

(f) Frame 1-Points (g) Frame 2-Points (h) Frame 3-Points (i) Frame 4-Points (j) Frame 5-Points

Figure 1. Frames from fourth recording of the sign "happiness".

The signs were captured using a RGB-D sensor (Microsoft Kinect) and processed by nuiCaptureAnalyze software. In the processing stage, each recording's images (Figures 1a to 1e) and xy-coordinates of 121 points located across the face were obtained (Figures 1f to 1j). This work focuses on the 121 points, which served as base descriptors for each facial expression.

Each sign in the original dataset was mapped to one of the labels as presented in Table 1, taking into consideration the closest fundamental expression or the most prominent facial parameter [Capovilla 2017]:

Table 1. Sign mapping

Sign	New label	Sign	New label
To calm down	Neutral	Happiness	Happy
To accuse	Angry	Skinny	Sucked cheeks
To annihilate	Angry	Lucky	Neutral
In love	Happy	Surprised	Surprised
To fatten	Inflated cheeks	Angry	Angry

3. Feature Extraction

In [Rezende et al. 2016], classification was performed on four different feature vectors composed by the 121 facial points:

- Utilizing raw data;
- Performing z-normalization on (x,y) coordinates separately, for each recording of each signal;
- Normalizing each xy-coordinate of each signal's recordings in relation to the centroid of the first corresponding frame;
- Normalizing each xy-coordinate of each signal's recordings in relation to the centroid of its current frame.

Due to data being highly dimensional in all aforementioned experiments (1210 features $\times$ 100 samples), this work proposes two novel approaches aiming to reduce feature space, discussed in the next subsections.

3.1. Points Selected by Inspection

Pairs of points were selected through inspection of video frames, tentatively capturing prominent characteristics of facial expressions. Each pair and its respective description is shown on Table 2 and the selected points can be seen in Figure 2.

Table 2. Selected pairs

Points	Description
6 and 3	Nose tip to mid supraorbital ridge
6 and 11	Nose tip to mid chin
8 and 9	Mouth opening height
20 and 25	Right eye opening
49 and 16	Outer eyebrows corners (left and right)
50 and 17	Eyebrow upper midpoint (left and right)
53 and 58	Left eye opening
65 and 32	Mouth opening width
91 and 92	Nasolabial folds (left and right)

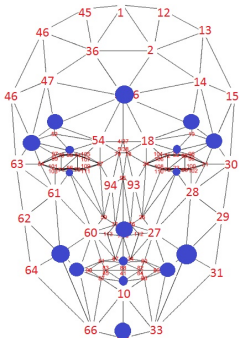


Figure 2. Selected points

Data dimension was reduced to 170 features (xy-coordinates of 17 points, 5 frames) $\times$ 100 samples. The resulting feature vector is as follows:

$$\underbrace{x_{i,j}}_{\text{signal i, recording j}} = \left[\underbrace{(s_x, s_y)_{1,1}\ (s_x, s_y)_{2,1}\ \ldots\ (s_x, s_y)_{s,f}}_{\text{point s coordinates, frame f}} \right] \tag{1}$$

3.2. Displacement Ranking

The original data set was processed following the steps bellow:

1. For each recording, a cumulative measure of displacement based on the Euclidean distance of a point in each consecutive frame was calculated;
2. Points were ranked within each signal, following the highest displacement value;

3. The first decile of each recording's rank was sampled, yielding 12 points with highest displacement per recording;
4. For each signal, a mode the first decile was determined;
5. Each signal's recordings were re-sampled from the original data with respect to each signal's mode. The new data set's dimensions are 120 features (xy-coordinates of 12 points, 5 frames) $\times$ 100 observations.

The resulting feature vector is as follows:

$$\underbrace{x_{i,j}}_{\text{signal i, recording j}} = \left[\underbrace{(p_x, p_y)_{1,1}\ (p_x, p_y)_{2,1}\ \ldots\ (p_x, p_y)_{p,f}}_{\text{point p coordinates, frame f}} \right] \tag{2}$$

4. Experimental Analysis

Three experiments were formulated to evaluate the proposed approaches:

1. Classification utilizing all 121 points;
2. Classification on the feature vector consisting of points selected by inspection of the video frames;
3. Classification on the reduced data set created through the displacement ranking.

All experiments were performed utilizing three classifiers, for means of comparison: Random Forest [Breiman 2001], Support Vector Machines [Cortes and Vapnik 1995] and k-Nearest Neighbors [Patrick and Fischer 1970]. The general structure of the solution is as seen on Algorithm 1. For each classifier, 30 models were constructed in each experiment, resulting in 90 models per experiment. Further performance measures were derived from predicted values and discussed later in this section.

Algorithm 1: Sign Classification

input : sign samples
output: predicted expression

maxIt $i \leftarrow$ *30*;
for $i \leftarrow 1$ **to** $maxIt$ **do**
    randomization of samples;
    train $i \leftarrow$ *80% of each class*;
    test $i \leftarrow$ *20% of each class*;
    parameters $\leftarrow$ *k-fold cross-validation*;
    model $\leftarrow$ *classifier(train* i*, parameters)*;
    predictions $i \leftarrow$ *model(test* i*)*;
end

4.1. Random Forest

Breiman's Random Forest (RF, [Breiman 2001]) is a powerful ensemble approach based on bootstrap aggregation of multiple decision trees, and is widely utilized in applications in which the number of features exceeds the number of observations. The algorithm differentiates from other decision tree methods in the way that at each node

split in the learning process, the feature space is randomly sampled with replacement. A final prediction is obtained by aggregating all constructed trees through majority voting [Boulesteix et al. 2012].

The Random Forest algorithm has several tuning parameters, of which most show high dependency on the data set [Ließ et al. 2012]. Hence, one of the practical challenges when using RF is parameterization. In this work, Random Forests parameters are selected through randomized search stratified 3-fold cross validation, obtaining 500 different settings [Pedregosa et al. 2011]. The search space was limited to according to the values shown in Table 3.

Table 3. Random Forest parameters

Parameter description	Possible values
Number of trees	[800, 2000], step size = 10
Number of features (p) considered at each split	$\{\log_2 p, \sqrt{p}, 0.3p\}$
Maximum depth of a tree	{[10, 80], step size = 10; or None}
Minimum number of samples required to split a node	{3, 5, 7};
Minimum number of samples required at each leaf-node	{2, 3, 4}
Whether bootstrapping occurs when constructing trees	{'True', 'False'}
Split quality measure	{'Gini', 'Entropy'}

4.2. Support Vector Machine

The SVM (Support Vector Machine) classifier presented by [Cortes and Vapnik 1995] is currently considered the state-of-art in classification and regression problems [Zhang et al. 2017]. The SVM algorithm finds points that make up support vectors, which, in turn, compose a hyperplane that optimizes the distance between the classes, serving as a decision boundary. This boundary is obtained using training data, and is applied to classify the test data.

The SVM solves binary classification problems. However, LIBRAS recognition is multiclass problem. Package e1071's implementation of SVM [Meyer and Wien 2001] uses a one-against-one technique for multiclass problems, thus, it was selected to be utilized in this work. In addition, package e1071 has tools to perform automatic search, by cross-validation, of the cost parameters C and gamma(γ), relative to the separation surface of the classes. [Hsu et al. 2016] advises that the parameter C to vary from 2^{-5} to 2^{15}, and γ, from 2^{-15} to 2^{3}. In relation to the kernel, Radial Base Function (RBF) was used, chosen according to [Hsu et al. 2016], after taking into consideration the cases tested in this work.

4.3. k-NN

The use of k-NN [Patrick and Fischer 1970] for classification problems is well established in the literature. Due to the classifier's efficiency in terms of running time

[Zhang et al. 2017], it was included in this work for comparison with SVM's and RF's results. To determine the class of a sample m not belonging to the training set, the k-NN classifier looks for k elements of the training set that are closest to m, and assigns its class based on which class represents the majority of these selected k elements.

With the selected training data, cross-validation was used to find the value for k that provided the highest accuracy rate (k_{best}). Thus, the training data were divided into 5-folds of the same size and 5 cross-validation iterations were performed applying the leave-one-out technique.

4.4. Results and Discussion

After applying the procedure shown in Algorithm 1, accuracy results were obtained for each classifier. Results were statistically analyzed utilizing ANOVA, Shapiro-Wilk and Fligner-Killeen tests [Elliott and Woodward 2007], and can be seen in Figure 3.

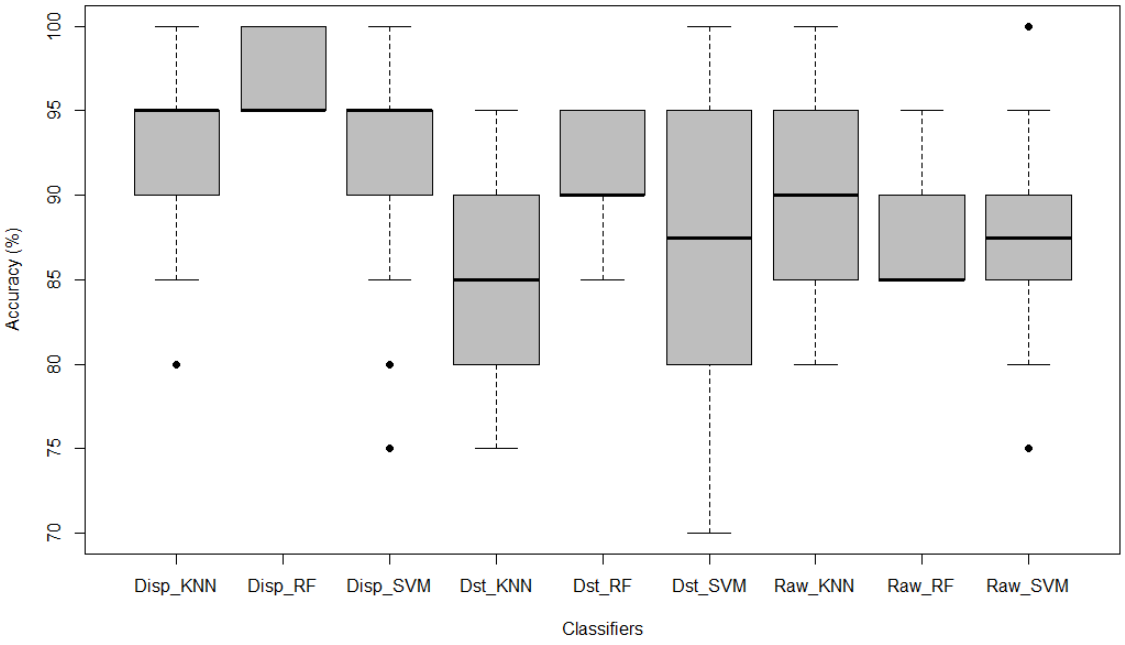


Figure 3. Comparison of results for each classifier

For classification with raw data, k-NN obtained the best average accuracy (89.33%). However, classifiers' overall performance are comparable. Utilizing points selected through inspection, Random Forest obtained the best average accuracy (91.33%) and achieved best overall performance within the experiment. SVM and k-NN presented similar performances. Finally, in the experiment with displacement ranking data, Random Forest presented the highest average accuracy (96.67%) and performed similarly to SVM, both surpassing k-NN's results.

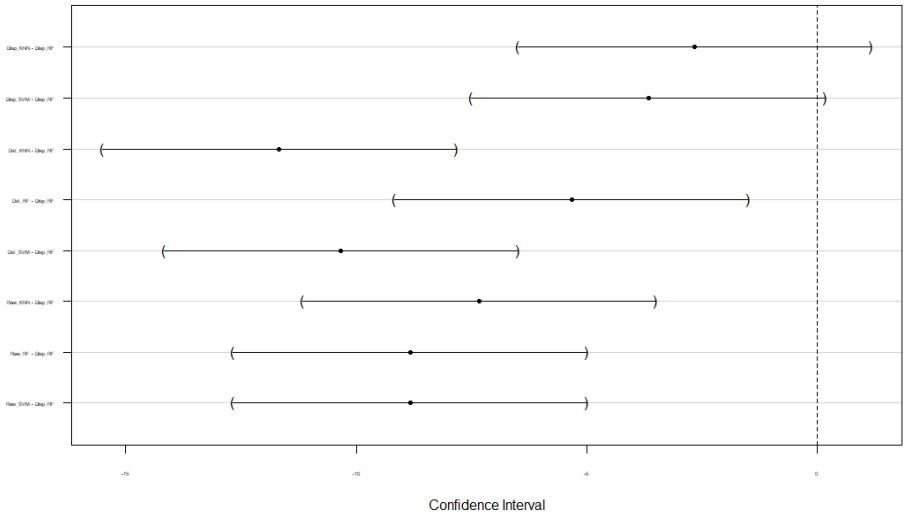


Figure 4. One-Against-All confidence interval of Random Forest + Displacement Ranking

Through an all-against-all analysis, as shown in Figure 4, the combination of displacement ranking and either Random Forest or SVM classifiers has shown the best overall results. Class-wise results for Random Forest and displacement ranking, seen in Table 4, show that the expression "Inflated Cheeks" had the lowest accuracy results, and was misclassified as "Happy".

Table 4. Aggregated confusion matrix for Random Forest + Displacement Ranking

Predicted	Angry	Happy	Surprised	Neutral	Sucked Cheeks	Inflated Cheeks	All
Angry	180	0	0	0	0	0	180
Happy	0	100	0	0	0	20	120
Surprised	0	0	60	0	0	0	60
Neutral	0	0	0	120	0	0	120
Sucked Cheeks	0	0	0	0	60	0	60
Inflated Cheeks	0	0	0	0	0	60	60
All	180	100	60	120	60	80	600

The work presented in this paper shows the following advancements when compared to [Rezende et al. 2016]: (i) validates Random Forest as an scalable alternative to SVM and k-NN; (ii) corroborates a new, generalizable approach to LIBRAS recognition, that can be combined to [Almeida et al. 2014] constituting a hollistic method to SLR.

5. Conclusion

This study proposed a new approach to LIBRAS recognition. In contrast to works presented in the literature, facial parameters were analyzed and classification was performed identifying basic elements that make up the structure of the language. Results validate the approach and introduce Random Forest as a good candidate for learning tasks.

Since non-manual configurations may be shared by different signs - or may not be used at all - future works addressing both manual and non-manual parameters are expected to deliver a holistic, and more precise, solution to SLR.

It is of fundamental importance that Computational Intelligence minimizes communication barriers and facilitates communication between those who have hearing impairments with those who do not. LIBRAS is not a compulsory component of the basic school curriculum at present, therefore sign language literacy level is low, making it hard for the deaf to communicate with the majority of the population.

6. Acknowledgments

This work has been supported by the Brazilian agency CAPES.

References

Abdullah, M. F. A., Sayeed, M. S., Muthu, K. S., Bashier, H. K., Azman, A., and Ibrahim, S. Z. (2014). Face recognition with symmetric local graph structure (SLGS). *Expert Systems with Applications*, 41(14):6131–6137.

Almeida, S. G. M., Guimarães, F. G., and Ramírez, J. A. (2014). Feature extraction in brazilian sign language recognition based on phonological structure and using RGB-d sensors. *Expert Systems with Applications*, 41(16):7259–7271.

Boulesteix, A.-L., Janitza, S., Kruppa, J., and König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(6):493–507.

Brasil (2002). Lei n^o 10.436, de 24 de abril de 2002.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Capovilla, F. C. (2017). *Dicionário da Língua de Sinais do Brasil. A Libras em Suas Mãos - 3 Volumes*. Edusp.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

de Assis Silva, C. A. (2012). Igreja católica e surdez: território, associação e representação política. *Religião & Sociedade*, 32(1):13–38.

de Paula Neto, F. M., Cambuim, L. F., Macieira, R. M., Ludermir, T. B., Zanchettin, C., and Barros, E. N. (2015). Extreme learning machine for real time recognition of brazilian sign language. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE.

Dias, D. B., Madeo, R. C. B., Rocha, T., Biscaro, H. H., and Peres, S. M. (2009). Hand movement recognition for brazilian sign language: A study using distance-based neural networks. In *2009 International Joint Conference on Neural Networks*. IEEE.

Du, S., Tao, Y., and Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences*, 111(15):E1454–E1462.

Elliott, A. and Woodward, W. (2007). *Statistical Analysis Quick Reference Guidebook*. SAGE Publications, Inc.

Escobedo-Cardenas, E. and Camara-Chavez, G. (2015). A robust gesture recognition using hand local data and skeleton trajectory. In *2015 IEEE International Conference on Image Processing (ICIP)*. IEEE.

Filho, C. F. F. C., de Souza, R. S., dos Santos, J. R., dos Santos, B. L., and Costa, M. G. F. (2017). A fully automatic method for recognizing hand configurations of brazilian sign language. *Research on Biomedical Engineering*, 33(1):78–89.

Freitas, F. A., Peres, S. M., Lima, C. A. M., and Barbosa, F. V. (2017). Grammatical facial expression recognition in sign language discourse: a study at the syntax level. *Information Systems Frontiers*, 19(6):1243–1259.

Genuer, R., Poggi, J.-M., Tuleau-Malot, C., and Villa-Vialaneix, N. (2017). Random forests for big data. *Big Data Research*, 9:28–46.

Gesser, A. (2009). *LIBRAS?: Que língua é essa?: crenças e preconceitos em torno da língua de sinais e da realidade surda*. Parábola Editorial, São Paulo.

Gross, R. (2005). Face databases. In S.Li, A., editor, *Handbook of Face Recognition*. Springer, New York.

Hinton, G., Deng, L., Yu, D., Dahl, G., rahman Mohamed, A., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., and Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6):82–97.

Hsu, C., Chang, C., and Lin, C. (2016). A practical guide to support vector classification.

Jung, H., Lee, S., Yim, J., Park, S., and Kim, J. (2015). Joint fine-tuning in deep neural networks for facial expression recognition. In *2015 IEEE International Conference on Computer Vision (ICCV)*. IEEE.

Laborit, E. (1998). *The cry of the gull*. Gallaudet University Press, Washington, DC.

Landar, H. and Stokoe, W. C. (1961). Sign language structure: An outline of the visual communication systems of the american deaf. *Language*, 37(2):269.

Ließ, M., Glaser, B., and Huwe, B. (2012). Uncertainty in the spatial prediction of soil texture. *Geoderma*, 170:70–79.

López, G., Quesada, L., and Guerrero, L. A. (2017). Alexa vs. siri vs. cortana vs. google assistant: A comparison of speech-based natural user interfaces. In *Advances in Intelligent Systems and Computing*, pages 241–250. Springer International Publishing.

Meyer, D. and Wien, T. U. (2001). Support vector machines. the interface to libsvm in package e1071. online-documentation of the package e1071 for r.

Pariwat, T. and Seresangtakul, P. (2017). Thai finger-spelling sign language recognition using global and local features with SVM. In *2017 9th International Conference on Knowledge and Smart Technology (KST)*. IEEE.

Patrick, E. and Fischer, F. (1970). A generalized k-nearest neighbor rule. *Information and control*, 16(2):128 – 152.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau,

D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Pigou, L., Dieleman, S., Kindermans, P.-J., and Schrauwen, B. (2015). Sign language recognition using convolutional neural networks. In *Computer Vision - ECCV 2014 Workshops*, pages 572–578. Springer International Publishing.

Porfirio, A. J., Wiggers, K. L., Oliveira, L. E., and Weingaertner, D. (2013). LIBRAS sign language hand configuration recognition based on 3d meshes. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE.

Pu, X., Fan, K., Chen, X., Ji, L., and Zhou, Z. (2015). Facial expression recognition from image sequences using twofold random forest classifier. *Neurocomputing*, 168:1173–1180.

Rao, G. A., Kishore, P. V. V., Sastry, A. S. C. S., Kumar, D. A., and Kumar, E. K. (2017). Selfie continuous sign language recognition with neural network classifier. In *Proceedings of 2nd International Conference on Micro-Electronics, Electromagnetics and Telecommunications*, pages 31–40. Springer Singapore.

Rezende, T. M., de Castro, C. L., and Almeida, S. G. M. (2016). An approach for brazilian sign language (bsl) recognition based on facial expression and k-nn classifier. In Fábio A. M. Cappabianco, Fábio A. Faria, J. A. and Körting, T. S., editors, *Conference on Graphics, Patterns and Images (SIBGRAPI '16)*. Sociedade Brasileira de Computação.

Uddin, M. A. and Chowdhury, S. A. (2016). Hand sign language recognition for bangla alphabet using support vector machine. In *2016 International Conference on Innovations in Science, Engineering and Technology (ICISET)*. IEEE.

Uddin, M. T. (2015). An ada-random forests based grammatical facial expressions recognition approach. In *2015 International Conference on Informatics, Electronics & Vision (ICIEV)*. IEEE.

Yang, H.-D. and Lee, S.-W. (2011). Combination of manual and non-manual features for sign language recognition based on conditional random field and active appearance model. In *2011 International Conference on Machine Learning and Cybernetics*. IEEE.

Yang, H.-D. and Lee, S.-W. (2013). Robust sign language recognition by combining manual and non-manual features based on conditional random field and support vector machine. *Pattern Recognition Letters*, 34(16):2051–2056.

Yu, Z. and Zhang, C. (2015). Image based static facial expression recognition with multiple deep network learning. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI '15*. ACM Press.

Zeng, Z., Pantic, M., Roisman, G., and Huang, T. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):39–58.

Zhang, C., Liu, C., Zhang, X., and Almpanidis, G. (2017). An up-to-date comparison of state-of-the-art classification algorithms. *Expert Systems with Applications*, 82:128–150.