International Journal of Humanities Social Science and Management (IJHSSM)
Volume 2, Issue 2, June 2022, pp: 05-11 www.ijhssm.org

Sign Language Recognition Using Deep Learning

Nithya M, *Computer Engineering(COM), Presidency University, Bangalore*
Bhagya K ,*Assistant professor, Department of Computer Science and Engineering Presidency University, Bangalore,*
Sandra Sai Rohith, *Computer Engineering(COM), Presidency University, Bangalore*
Devineni Yamini, *Computer Science and Engineering, Presidency University, Bangalore*
GantasalaSahithya, *Computer Science and Engineering, Presidency University, Bangalore*

Date of Submission: 24-06-2022 Date of Acceptance: 05-07-2022

ABSTRACT: Sign language is used by speech and hard of hearing people communicate, but normal people no longer understand sign language, causing a conversation gap between them. So, we need a system that closes the gap between normal people and people with disabilities. Sign languages are non-standard, and often, they vary from country to country. In this article, we used American Sign Language (ASL) to interpret the signs. The test we conducted showed an accuracy of 95% for the MNIST dataset. Here, the algorithm used is Convolution Neural Network (CNN). The deployed solution was trained to recognize all alphabets by deep learning.

KEYWORDS: Sign Language, Impaired people American Sign Language (ASL), Convolution Neural Network (CNN), MNIST dataset, Deep Learning

I. INTRODUCTION:

Speech impaired people face plenty of trouble in communicating with people around them. Their relations and near ones learn sign language to speak with them every day, aside from the family members other humans don't learn sign language because they don't use it every day. But when impaired people want to speak with normal people other than family members they don't understand. Society doesn't recognize how difficult it's miles for them every day to talk with someone who does no longer realize sign language.Sign language is a language in which verbal exchanges among humans are made using visually transmitting the sign styles every day. signing additionally includes facial features and frame language(body language), those are called non-manual signs. They play a major role in information correct which means understating the correct meaning of signs. Manual signs are the hand patterns and it carries the most information at the same time as non-manual signs are important for the explanation. Speech impaired people study signing so they can communicate with people around them and express their thoughts.

Sign language is different in each country. Special international locations have their signing, as an example, American signing (ASL), Argentinean signing (LSA), British signing (BSL), and Indian signing (ISL). Speech impaired people choose signing that is used in their particular area. to create the conversation among speech impaired people and regular people, an indication language recognition system is used which converts sign language into text by using some technologies.

Sign language recognition may be done through two techniques. the primary one is senor based technique, during this approach sensor-based wearable devices are used including gloves. This approach has two hazards, first one is that the devices are high priced and second those gadgets have to be carried everywhere otherwise the machine would no longer work. The second technique includes laptop vision-based methods. during this approach, the dataset is employed for recognition. This approach is more convenient to use as it does not need the user to put on any devices.

Many studies are executed on sign language recognition but most of the work is accomplished on American sign language (ASL). Different languages aren't explored as a good deal as ASL. ASL is a single-hanged signing, which makes it simpler to work with, whereas in Indian signing (ISL) some signs are carried out with one hand and some signs are carried out with the usage of both hands which makes it complicated to work with.

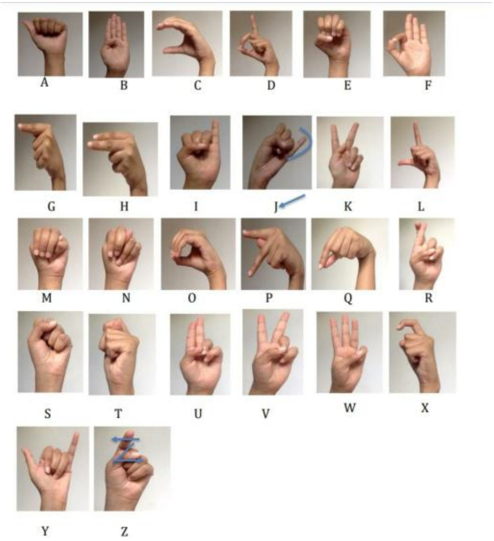


Figure 1: American Sign Language Alphabets

1.1 MOTIVATION

Communication is one of the number one necessity for survival in society. Speech impaired human beings communicate among themselves using signing but each day normal humans locate it is troublesome to acknowledge their language. monumental work has been administered on us linguistic communication name however Indian linguistic communication differs notably from us linguistic communication. ISL operates arms for communication whereas ASL uses one hand for communication. The use of every palm regularly outcomes within the protection of the abilities duet overlapping of palms. In addition to the present loss of datasets at a side of variance in signing with neighbourhood has resulted in restricted efforts in ISL gesture detection. Our undertaking targets assault the elemental step in bridging the communication hallow between ordinary human beings and speech impaired humans communicate quicker and easier with the outer world but also offer a boost in growing a self-maintaining machine for know-how and assisting them.

1.2 PROBLEM STATEMENT

Sign language makes use of a number of gestures so that it seems like motion language which includes a sequence of hands and finger motions. For one-of-a-type countries, there are distinct sign languages and hand gestures, additionally, it’s mentioned that some unknown words are translated with the aid of truly displaying gestures for every alphabet in the phrases. In addition, signing additionally includes precise gestures to every alphabet inside the English dictionary. This work focuses on the advent of a static signing translator and the usage of a Convolutional Neural Network.

1.3OBJECTIVES

The main objective of this piece of work is to contribute to the sector of linguistic communication recognition and laptop text translation. In our challenge, we specialize in static sign language hand gestures. This work focuses on hand gesture recognition, including 26 alphabets using deep neural networks. We have created a complex neural network classifier that will classify hand gestures into English alphabets. We labelled the neural network under proprietary configurations.

II. IMPLEMENTATION

2.1 Dataset

We have used MNSIT datasets and trained the model to achieve good accuracy.

2.1.1 ASL Alphabet

The data is collected in the form of a CSV file having 24 alphabets with labels and pixels that represent the various classes.
columns and has different pixels. The testing dataset consists of 7172 rows and 785 columns. There are 24 classes of English alphabet (A-Z) excluding J and Z..

2.2 Convolutional Neural Network (CNN)

Computer imagination and prescient may be a field of artificial intelligence that makes a specialty of troubles related to photographs and motion pictures. CNN combined with computer vision can perform complicated issues. Convolutional Neural Networks are deep neural networks that have a grid-like topology, for instance, pictures which is able to be displayed as a 2nd array of pixels. The CNN has 2 main phases particularly feature extraction and classification. And additionally a series of convolution and pooling operations ar performed to extract the options of the images . a totally connected layer in convolutional neural networks can perform as a classifier. among the last word layer, the likelihood of magnificence ar aiming to be anticipated. The foremost steps concerned in convolutional neural networks are:

- ➢ Convolution
- ➢ Relu
- ➢ Pooling
- ➢ Flatten
- ➢ Full connection

2.2.1 Convolution

Convolution is nothing more than a filter applied to an image to extract features. We will use proprietary filters to extract features such as edges and strong patterns in the image. Filters are usually randomly generated. What this convolution will is produce a filter of a precise size wherever 3x3 is that the default size. After expanding the filter, it starts to do element-by-element multiplication from the vertices (top) of the left corner (corner) of the image to the bottom right of the image. The results can also be extracted features.

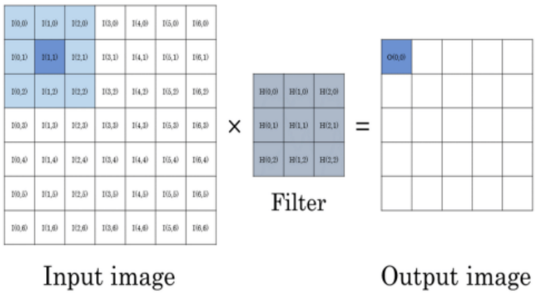

I(0,0) I(1,0) I(2,0) I(3,0) I(4,0) I(5,0) I(6,0)
I(0,1) I(1,1) I(2,1) I(3,1) I(4,1) I(5,1) I(6,1)
I(0,2) I(1,2) I(2,2) I(3,2) I(4,2) I(5,2) I(6,2)
I(0,3) I(1,3) I(2,3) I(3,3) I(4,3) I(5,3) I(6,3)
I(0,4) I(1,4) I(2,4) I(3,4) I(4,4) I(5,4) I(6,4)
I(0,5) I(1,5) I(2,5) I(3,5) I(4,5) I(5,5) I(6,5)
I(0,6) I(1,6) I(2,6) I(3,6) I(4,6) I(5,6) I(6,6)
×
H(0,0) H(1,0) H(2,0)
H(0,1) H(1,1) H(2,1)
H(0,2) H(1,2) H(2,2)
Filter
=
O(0,0)
Input image
Output image

Figure 2: Convolution

The length of the output matrix decreases as we generally tend to hold making use of the filters..
Size of latest matrix = (Size of antique matrix – clear out out size) + 1.
Convolution is a technique for extracting alternatives from an input image. It learns image alternatives by victimising small squares in the input file, preserving the abstraction relationship between pixels. Relu is usually the one who comes after it.

2.2.2 Relu:

In Relu All negative element values in the feature map are replaced with zero in this associate degree element-wise process. Its goal is to introduce nonlinearity into a convolutional network.

2.2.3 Max Pooling

It is choosing the utmost element price from the matrix.This methodology is useful for extracting the most important options or that ar highlighted within the image.It is additionally referred to as down-sampling that within the feature map, this replaces any negative element values however retains necessary information.

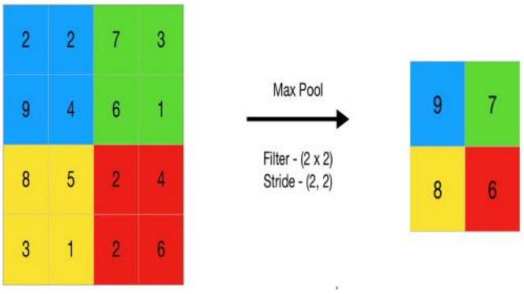


Figure 3: Max pooling

2.2.4 Flatten

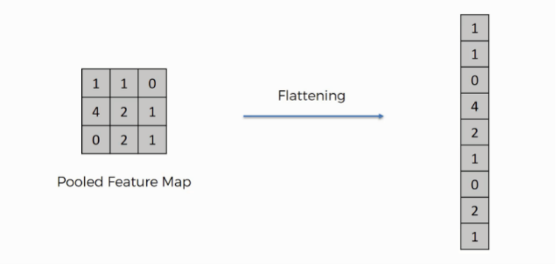


Figure 4: Flatten

The obtained resultant matrix square measure attending to be in multi-dimension. Flattening is changing the information into a 1-Dimensional array for inputting the layer to successive layer. we have a tendency to flatten the convolution layers to form one feature vector.

2.2.4 Full Connection

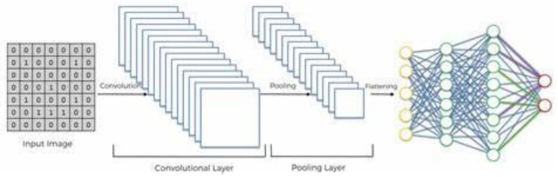


Figure 5: Full Connection

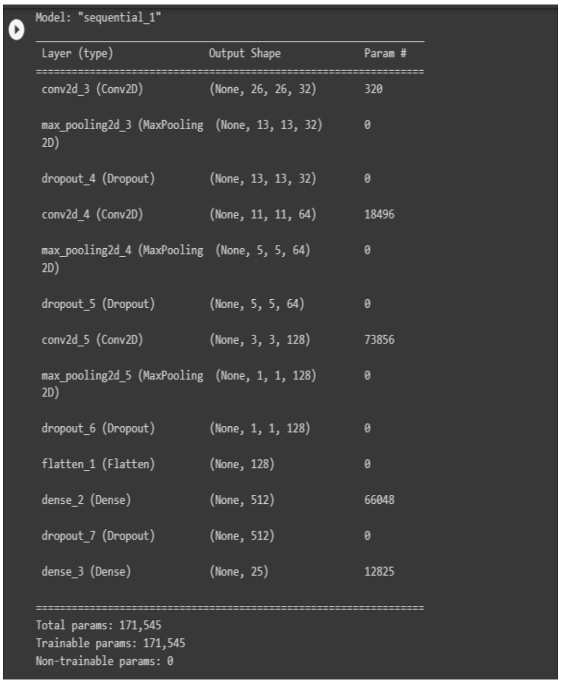
A full connected layer is simply a feed-forward neural network. All the operations area unit reaching to be performed and prediction is obtained. supported the bottom truth the loss are calculated and weights area unit updated victimization the gradient descent backpropagation algorithmic program. It's a multi-layer perceptron with a softmax function in the output layer. Its objective is to support coaching information by using the helpful options from previous layers for categorising the input image into various categories. A CNN model is created by combining all of these layers. The final layer is completely interconnected.

2.2.5 Pre-training a CNN model

Where the model is initially pre-trained on a dataset that is completely different from the first, the idea of Transfer learning is applied. In this way, the model accumulates data that can be transferred to other neural networks. The model's data is recorded in the form of "weights" at intervals and can be fed into another model. By layering fully-connected layers on top of the pre-trained model, it is commonly employed as a feature extractor. After loading the saved weights, the model is trained with the primary dataset.

TRANSFER LEARNING

We use a pre-trained model that has been trained on a massive dataset, and we transfer the weights that have been learned during many hours of coaching to several high-powered GPUs. For classification, we'll utilise the basic -V3 Model: Trained with a dataset of 1000 categories from the main ImageNet dataset, which was trained with over one million coaching images. The key difference between starting models and standard CNNs is that starting blocks are smaller. These methods entail concatenating the results of numerous filters on the same input tensor. Regular CNNs, on the other hand, do one convolution operation per tensor.



```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d_3 (Conv2D)           (None, 26, 26, 32)        320

 max_pooling2d_3 (MaxPooling  (None, 13, 13, 32)       0
 2D)

 dropout_4 (Dropout)         (None, 13, 13, 32)        0

 conv2d_4 (Conv2D)           (None, 11, 11, 64)        18496

 max_pooling2d_4 (MaxPooling  (None, 5, 5, 64)         0
 2D)

 dropout_5 (Dropout)         (None, 5, 5, 64)          0

 conv2d_5 (Conv2D)           (None, 3, 3, 128)         73856

 max_pooling2d_5 (MaxPooling  (None, 1, 1, 128)        0
 2D)

 dropout_6 (Dropout)         (None, 1, 1, 128)         0

 flatten_1 (Flatten)         (None, 128)               0

 dense_2 (Dense)             (None, 512)               66048

 dropout_7 (Dropout)         (None, 512)               0

 dense_3 (Dense)             (None, 25)                12825

=================================================================
Total params: 171,545
Trainable params: 171,545
Non-trainable params: 0
```

Figure 6: Model summary

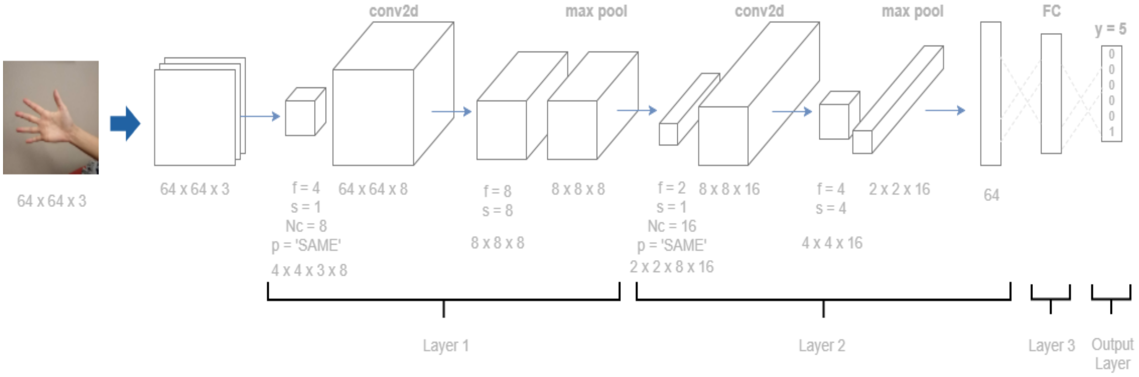


Figure 7 – A working model

III. PROPOSED METHOD:

In this paper, we are proposing a way that is useful for recognizing the signs through hand gestures employing a Convolution Neural Network (CNN) based mostly transfer learning algorithmic program from deep learning.Once after considering the dataset of various signs, we'll be training the dataset using a transfer learning algorithm and then the model can be used for the detection of the particular sign. The picture of the proposed method is below.

3.1 Advantages:

- ➢ Accurate classification
- ➢ Less complexity
- ➢ High performance

3.2 BLOCK DIAGRAM:

❖ Hand gesture Dataset:

We have collected the MNIST dataset for testing and training. The dataset is a CSV extension.

❖ Pre-Processing:

Pre-processing techniques are used to remove unwanted noise and improve the quality of an Input image.

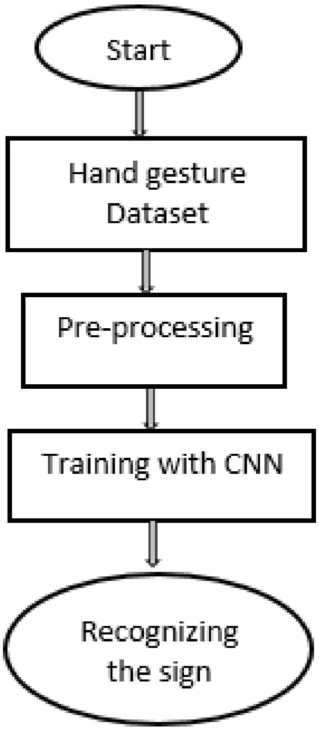


Figure 8: diagram of the proposed method

❖ Training with CNN:

We are training our dataset by using a deep neural network and the model we train is CNN.

❖ Recognizing the sign:

Finally, we are identifying the sign in the text format.

IV. Experiment Results:

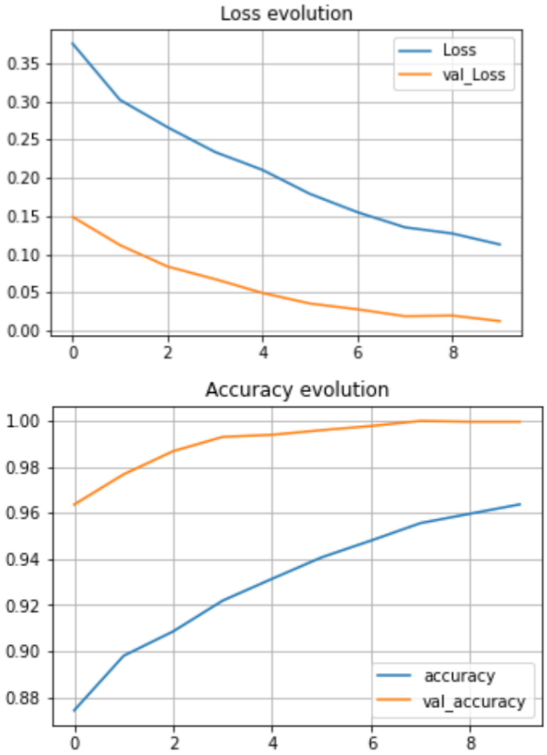


Figure 9: Training Graphs

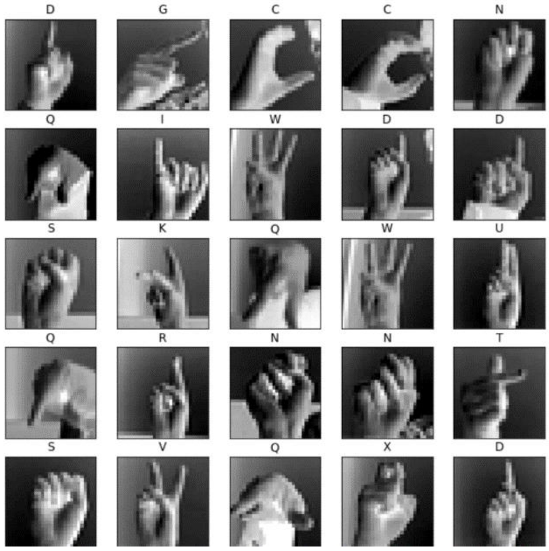
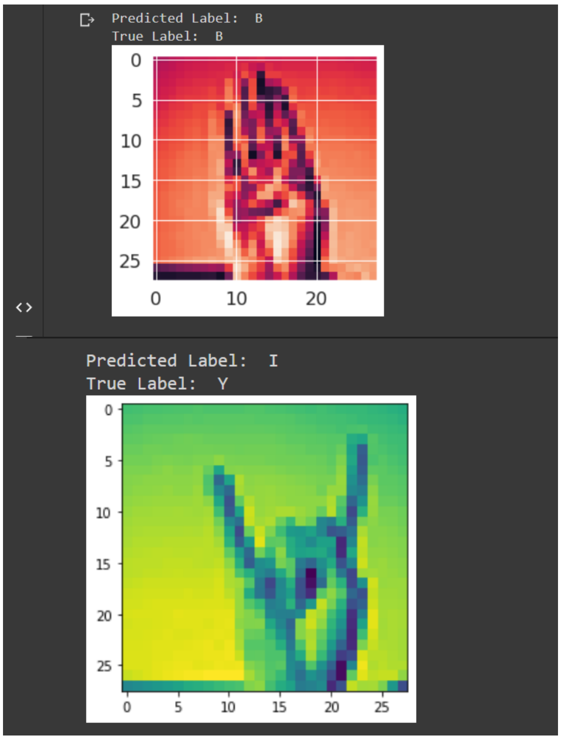


Figure 10: Model Prediction during training

We have tested the test dataset and we have achieved almost 94% of accuracy.



```
Predicted Label:  B
True Label:  B
```

```
Predicted Label:  I
True Label:  Y
```

Figure 11: Testing Model Output

V. Future Scope:

The studies examine the advantages to recognize 26 ASL alphabets using static hand gestures along with the natural lighting condition.

- We can develop a model for ISL word sentence level recognition. this could need a system which will sight changes regarding the temporal house.
- We can develop an entire product which will facilitate the speech and hard-of-hearing individuals and thereby scale back the communication gap.
- We will attempt to acknowledge signs that embrace motion. Moreover, we'll specialize in changing the sequence of gestures into text i.e. words and sentences, then changing it into speech that may be detected.

VI. Conclusion:

In conclusion, we have a tendency to be success and able to develop a sensible and understandable system that may acknowledge the signs and predict the true label. There are still many shortagesof those systems as we didn't use the camera for the development. We are sure it are oftenimplemented and optimized in the future.

References:

[1]. Wachs, J.P., H. Stern, and Y. Edan, Cluster Labeling and Parameter Estimation for the Automated Setup of a Hand-Gesture Recognition System. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2005. 35(6): p. 932-944.

[2]. Trigueiros, P., F. Ribeiro, and L.P. Reis. A Comparative Study of different image features for hand gesture machine learning. in 5th International Conference on Agents and Artificial Intelligence. 2013. Barcelona, Spain.

[3]. Trigueiros, P. and F. Ribeiro. Vision-based Hand WheelChair Control. in 12th International Conference on Autonomous Robot Systems and Competitions. 2012. Guimarães, Portugal.

[4]. Wang, C.-C. and K.-C. Wang. Hand Posture Recognition Using Adaboost with SIFT for Human-Robot Interaction. in Proceedings of the International Conference on Advanced Robotics 2008. Jeju, Korea.

[5]. Conseil, S., S. Bourenname, and L. Martin, Comparison of Fourier Descriptors and Hu Moments for Hand Posture Recognition, in 15th European Signal Processing Conference (EUSIPCO)2007: Poznan, Poland. p. 1960-1964.

[6]. Barczak, A.L.C., et al., Analysis of Feature Invariance and Discrimination for Hand Images: Fourier Descriptors versus Moment Invariants, in International Conference Image and Vision Computing2011: New Zeland.

[7]. Bourennane, S. and C. Fossati, Comparison of shape descriptors for hand posture recognition in video. Signal, Image and Video Processing, 2010. 6(1): p. 147-157.

[8]. Triesch, J. and C.v.d. Malsburg. Robust Classification of Hand Postures against Complex Backgrounds. in International Conference on Automatic Face and Gesture Recognition. 1996. Killington, Vermont, USA.

[9]. Huynh, D.Q. Evaluation of Three Local Descriptors on Low-Resolution Images for Robot Navigation. in 24th International Conference Image and Vision Computing. 2009. Wellington, New Zealand.

[10]. Fang, Y., et al. Hand Posture Recognition with Co-Training. in 19th International Conference on Pattern Recognition. 2008. Tampa, FL, USA.

[11]. Blum, A. and T. Mitchell Combining labeled and unlabeled data with co-training, in Proceedings of the eleventh annual conference on Computational learning theory1998, ACM: Madison, Wisconsin, United States. p. 92-100.

[12]. Tara, R.Y., P.I. Santosa, and T.B. Adji, Sign Language Recognition in Robot Teleoperation using Centroid Distance Fourier Descriptors.International Journal of Computer Applications, 2012. 48(2).

[13]. Faria, B.M., N. Lau, and L.P. Reis. Classification of Facial Expressions Using Data Mining and machine Learning Algorithms. in 4ª Conferência Ibérica de Sistemas e Tecnologias de Informação. 2009. Póvoa de Varim, Portugal.

[14]. Faria, B.M., L.P. Reis, and N. Lau. Cerebral Palsy EEG Signals Classification: Facial Expressions and Thoughts for Driving an Intelligent Wheelchair. in Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on. 2012.

[15]. Gillian, N.E., Gesture Recognition for Musician Computer Interaction, in Music Department2011, Faculty of Arts, Humanities and Social Sciences: Belfast. p. 206.

[16]. Faria, B.M., et al., Machine Learning Algorithms Applied to the Classification of Robotic Soccer Formations and Opponent Teams, in IEEE Conference on Cybernetics and Intelligent Systems (CIS)2010: Singapore. p. 344 - 349

[17]. Mannini, A. and A.M. Sabatini, Machine learning methods for classifying human physical activity from on-body accelerometers.Sensors (Basel), 2010. 10(2): p. 1154-75.

[18]. Maldonado-Báscon, S., et al., Road-Sign detection and Recognition Based on Support Vector Machines, in IEEE Transactions on Intelligent Transportation Systems2007. p. 264-278