THESIS

Automatic Method for the Recognition of Hand Gestures for the Categorization of Vowels and Numbers in Colombian Sign Language

Tutor: Roger Guzman Avendaño

Member 1: Gabriel Jiménez Forero — Código: 625455
Member 2: Evert Esneyder Moreno Mosquera — Código: 625488

SAPIENTIA AEDIFICAVIT SIBI DOMUM
UNIVERSIDAD CATÓLICA de Colombia
Vigilada Mineducación

Outline

- Problem statement
- Objectives
- Methodology
- Results
- Discussions
- Conclusion
- Products
- Future works
- Bibliography

Problem Statement (1 / 2)

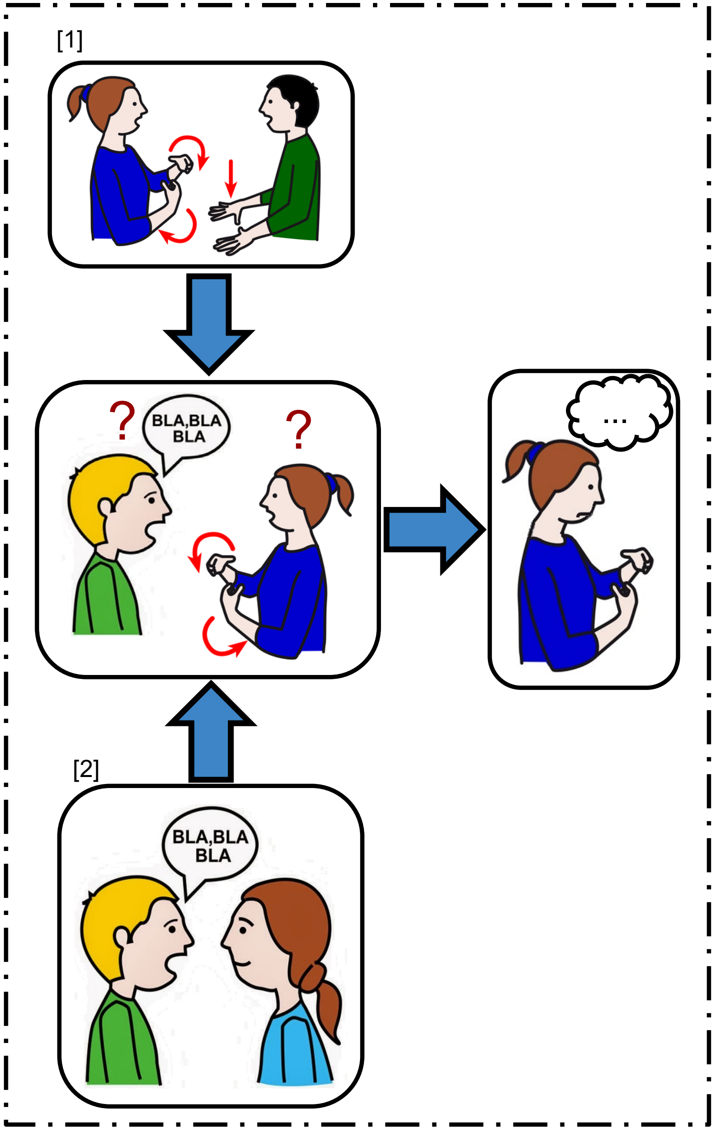


Figure 1: Problem communication of deaf

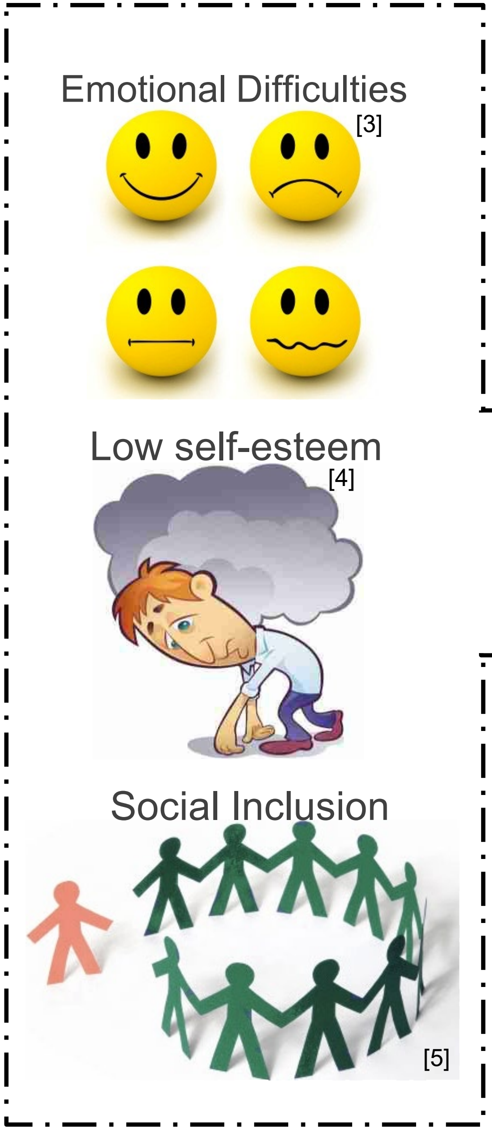


Figure 2: Implications of being deaf

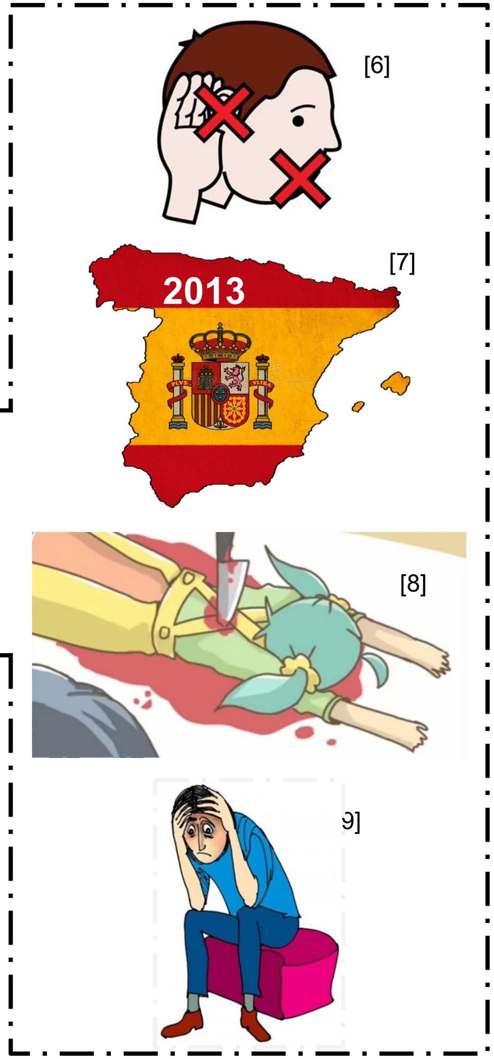


Figure 3: Case

Problem Statement (2 / 2)

[1]

Figure 4: Distractions

[2]

Figure 5: Subjectivity

[3]

Figure 6: Efficiency

Objectives

General Objectives

To implement an Automatic Classification method using Machine Learning to identify images of vowel and number gestures in Colombian Sign Language

Specific Objectives

- To build a data set with pictures of vowels and numbers from 0 to 5 in Colombian Sign Language with their respective labels.
- To develop a strategy processing, sampling and classification to identify Hand Gestures Colombian Sign Language using Machine Learning algorithms.
- To compare three methods of image classification, using performance measures such as precision, recall and F1 Score.

Workflow Hand Gesture Recognition

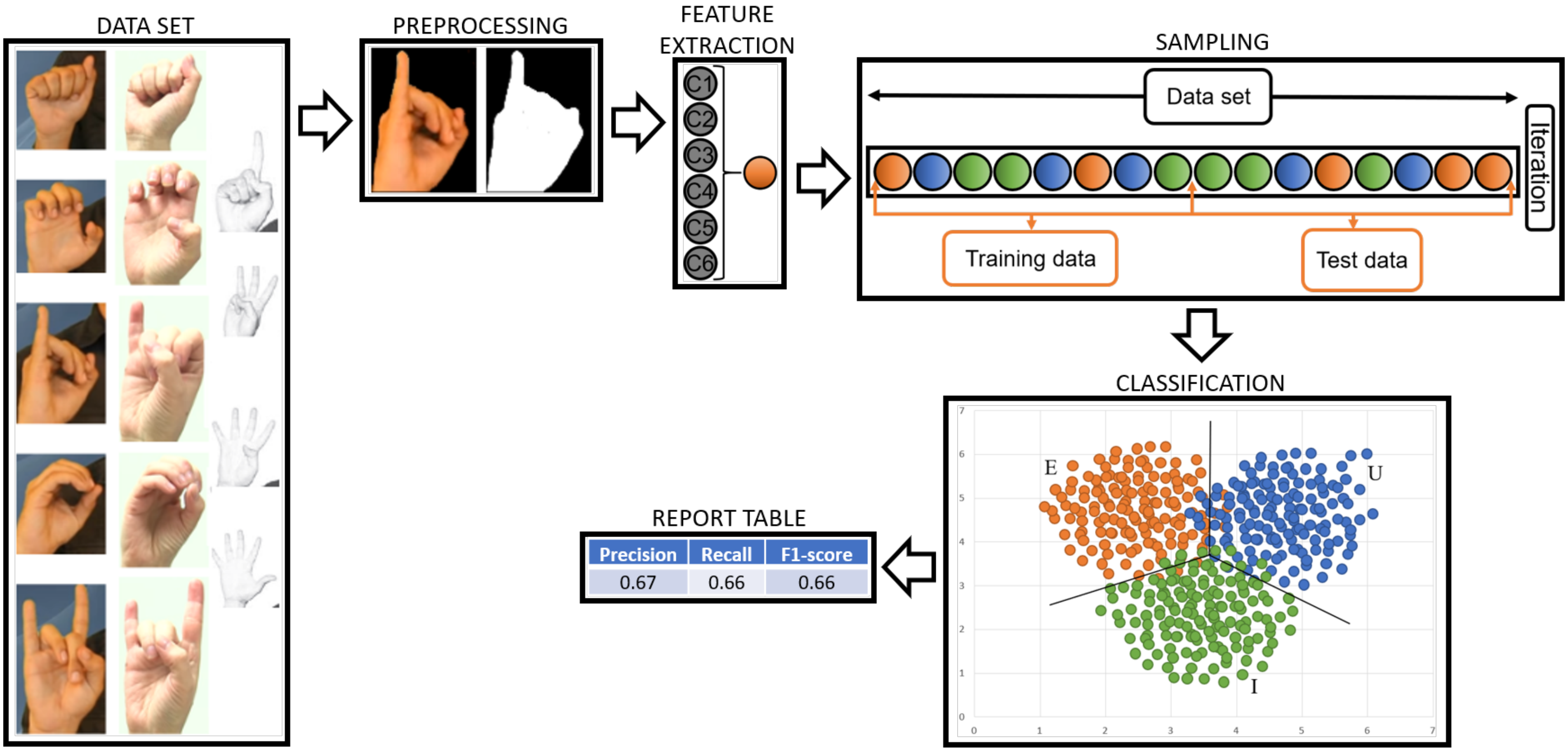


Figure 7: Workflow Hand Gesture Recognition

Dataset Description

Figure 8: Take pictures of hands

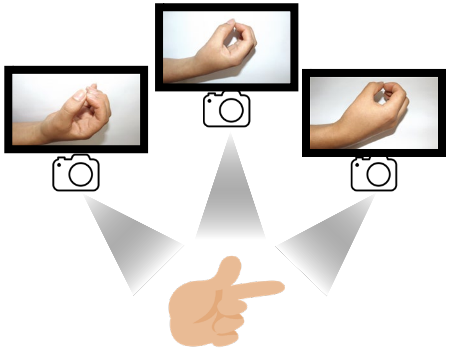


Figure 9: Build Data Set

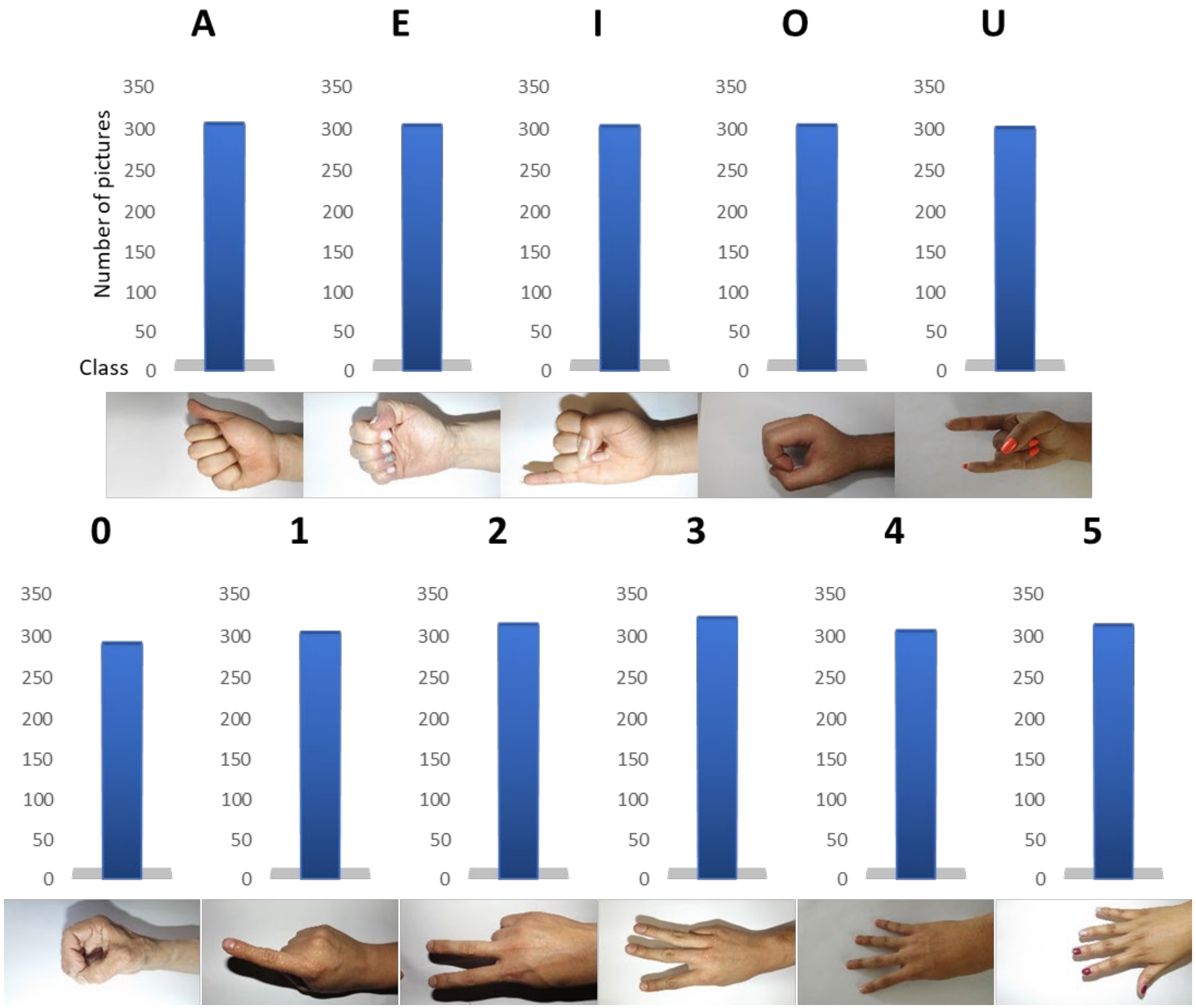


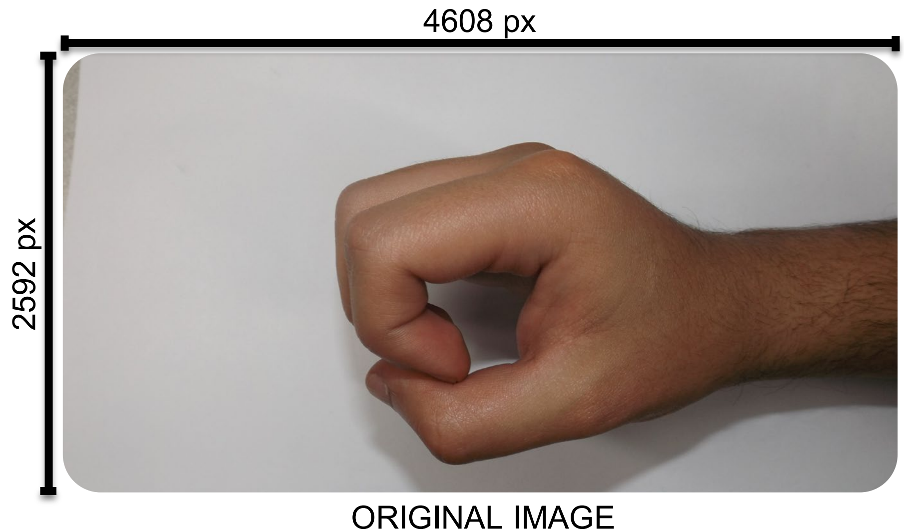
Figure 10: Dataset

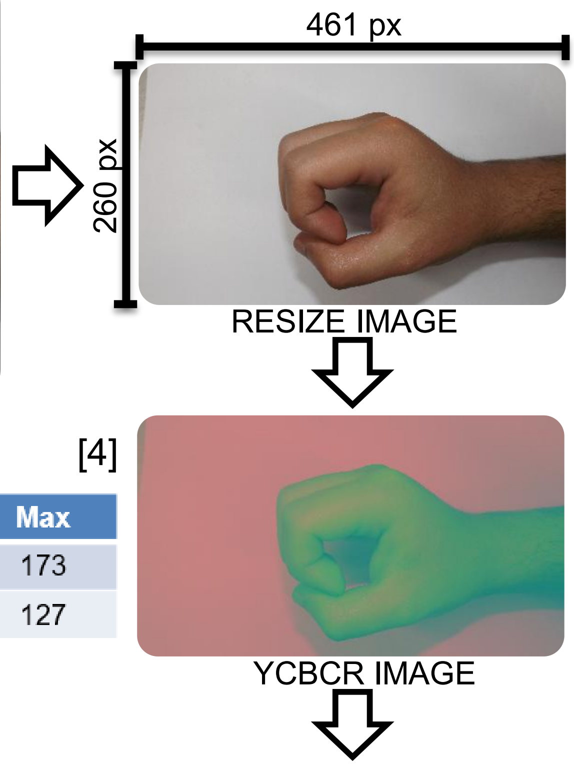
Total dataset: 3324 photos.

UNIVERSIDAD CATÓLICA de Colombia

Vigilada Mineducación

Workflow Preprocessing





[4]

Min	Chanel	Max
133	< Cr >	173
77	< Cb >	127

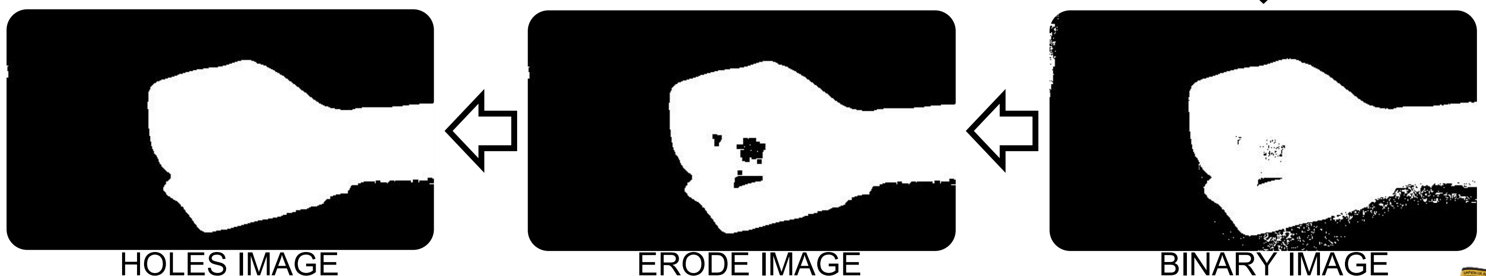


Figure 11: Workflow Preprocessing

Feature Extraction (1/4)
(Hu Moments)

A continuation presents the general formula to get the Hu moments:

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q f(x,y)\,dxdy$$

Where p and q = 0, 1, 2,…

- The zero-order moment where (p = q = 0) matches the image area.
- The moments of order one (p = 0, q = 1 and p = 1, q = 0), together with the zero order, determine the center of gravity of the images
- The central moments can be made invariant to the translations. This means that we refer to the center of gravity of the object
- The second moment and the third moment, derive from a set of seven invariant moments to translations, rotations and scale changes

Feature Extraction (2/4)
(Histogram of Oriented Gradients (HOG))

The gradient vector of the pixel's (x,y):

$$\nabla f(x,y) = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\partial_f}{\partial_x} \\ \frac{\partial_f}{\partial_y} \end{bmatrix} = \begin{bmatrix} f(x,y+1) & -\ f(x,y-1) \\ f(x+1,y) & -\ f(x-1,y) \end{bmatrix}$$

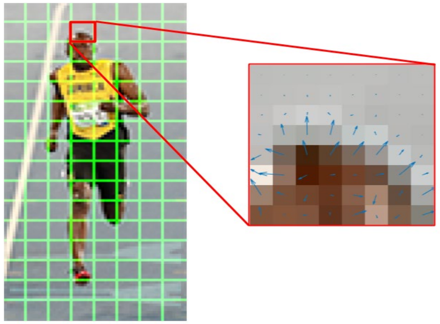


Figure 12: Getting pixels

Magnitude is the L2−norm of the vector

$$g = \sqrt{g_x^2 + g_y^2}$$

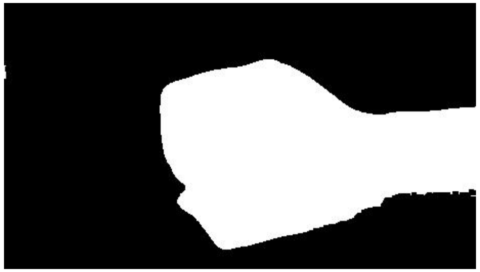


Figure 13: Segmented image

Direction is the arctangent of the ratio between the partial derivatives on two directions

$$\theta = \arctan(\frac{g_y}{g_x})$$

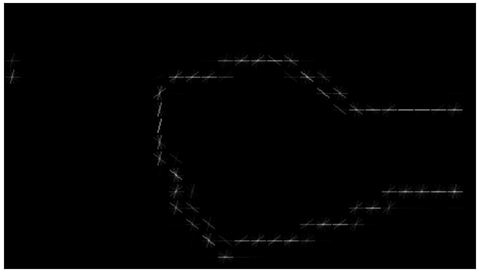


Figure 14: Image with HOG

Feature Extraction (3/4)

(Morphological features)

Área

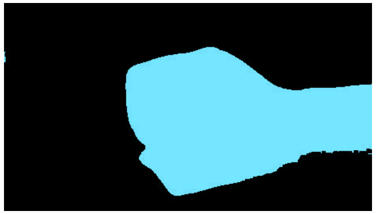


Figure 15: Hand Area

Centroid

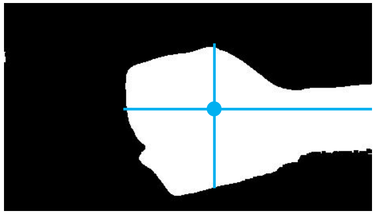


Figure 16: Hand centroid

Convex Hull

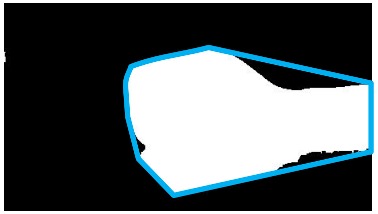


Figure 17: Hand Convex Hull

Aspect Ratio

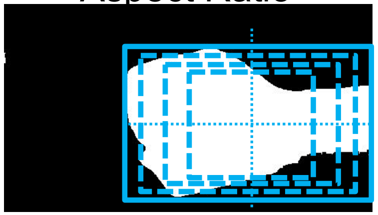


Figure 18: Hand Aspect ratio

$$\text{Aspect Ratio} = \frac{Width}{Height}$$

Eccentricity

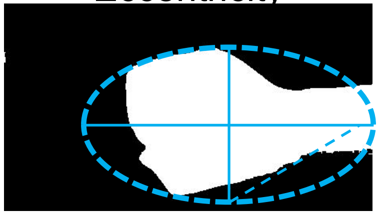


Figure 19: Hand Eccentricity

$$\text{Eccentricity} = \frac{Object\ Area}{Bounding\ Rectangle\ Area}$$

Solidity

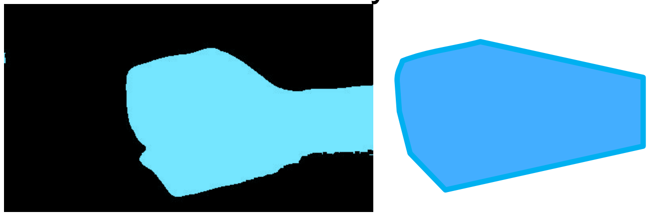


Figure 20: Hand Solidity

$$\text{Solidity} = \frac{Contour\ Area}{Convex\ Hull\ Area}$$

Equivalent Diameter

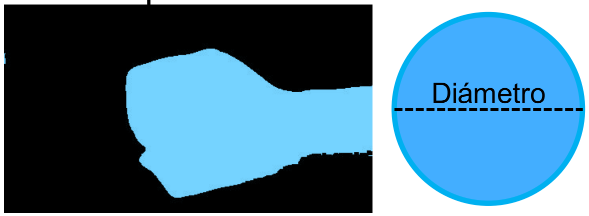


Figure 21: Hand Equivalent diameter

$$\text{Equivalent Diameter} = \sqrt{\frac{4 * Contour\ Area}{\pi}}$$

Feature Extraction (4/4)
(Elliptic Fourier)

Fourier's elliptical expansion of the coordinate sequence in x and y:

$$x(t) = \sum_{n=1}^{N} [A_n \cos\left(\frac{2\pi n t}{T}\right) + B_n \sin\left(\frac{2\pi n t}{T}\right)]$$

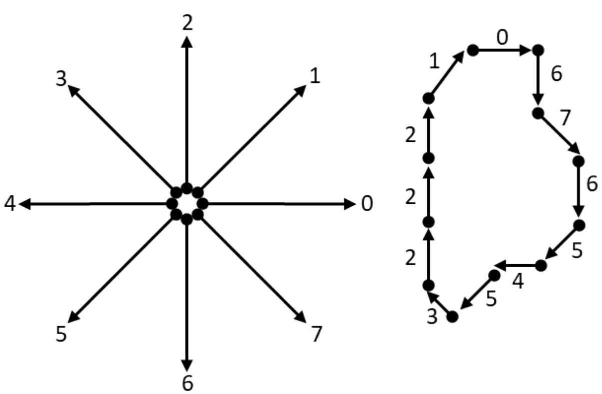
$$y(t) = \sum_{n=1}^{N} [C_n \cos\left(\frac{2\pi n t}{T}\right) + D_n \sin\left(\frac{2\pi n t}{T}\right)]$$

- n = The harmonic number.
- N = The maximun harmonic number.
- t = Displacement along outline.
- T = Total displacement.

$$A_n = C_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^{k} \frac{\Delta x_p}{\Delta t_p} [\cos\left(\frac{2\pi n t_p}{T}\right) - \cos\left(\frac{2\pi n t_{p-1}}{T}\right)]$$

$$B_n = D_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^{k} \frac{\Delta x_p}{\Delta t_p} [\sin\left(\frac{2\pi n t_p}{T}\right) - \sin\left(\frac{2\pi n t_{p-1}}{T}\right)]$$

Where:
- k = the total number of steps around the outline
- n = the harmonic number
- Δx = the displacement along the x axis between point p and p+1
- Δt = the length of the step between point p and p+1
- tp = accumulated length of step segments at point p
- T = sum of lengths of all steps around outline



Example: 2,2,2,1,0,6,7,6,5,4,5,3

Figure 22: Chain code

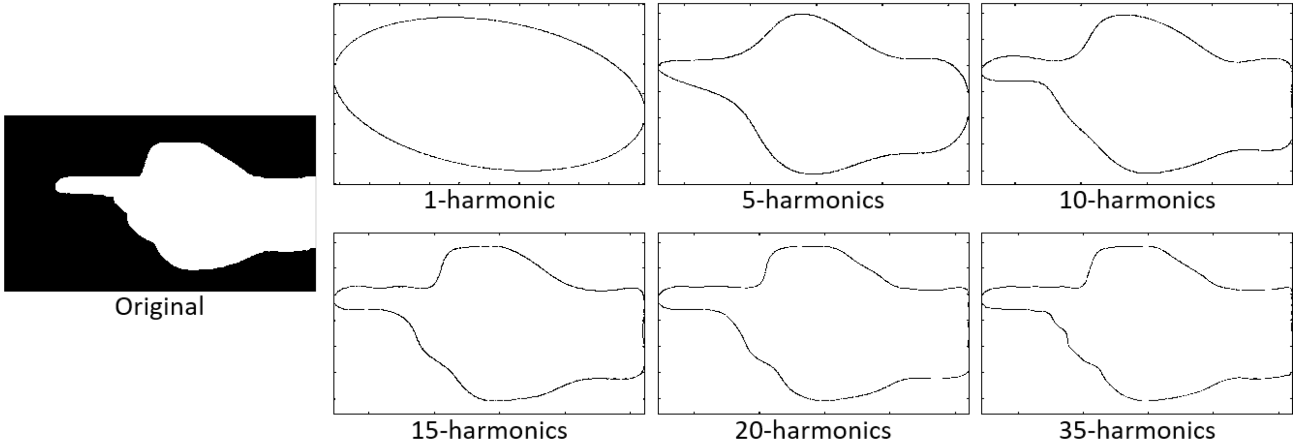


Figure 23: Hand with harmonics

UNIVERSIDAD CATÓLICA de Colombia
Vigilada Mineducación

Sampling

(Cross Validation K-folds)

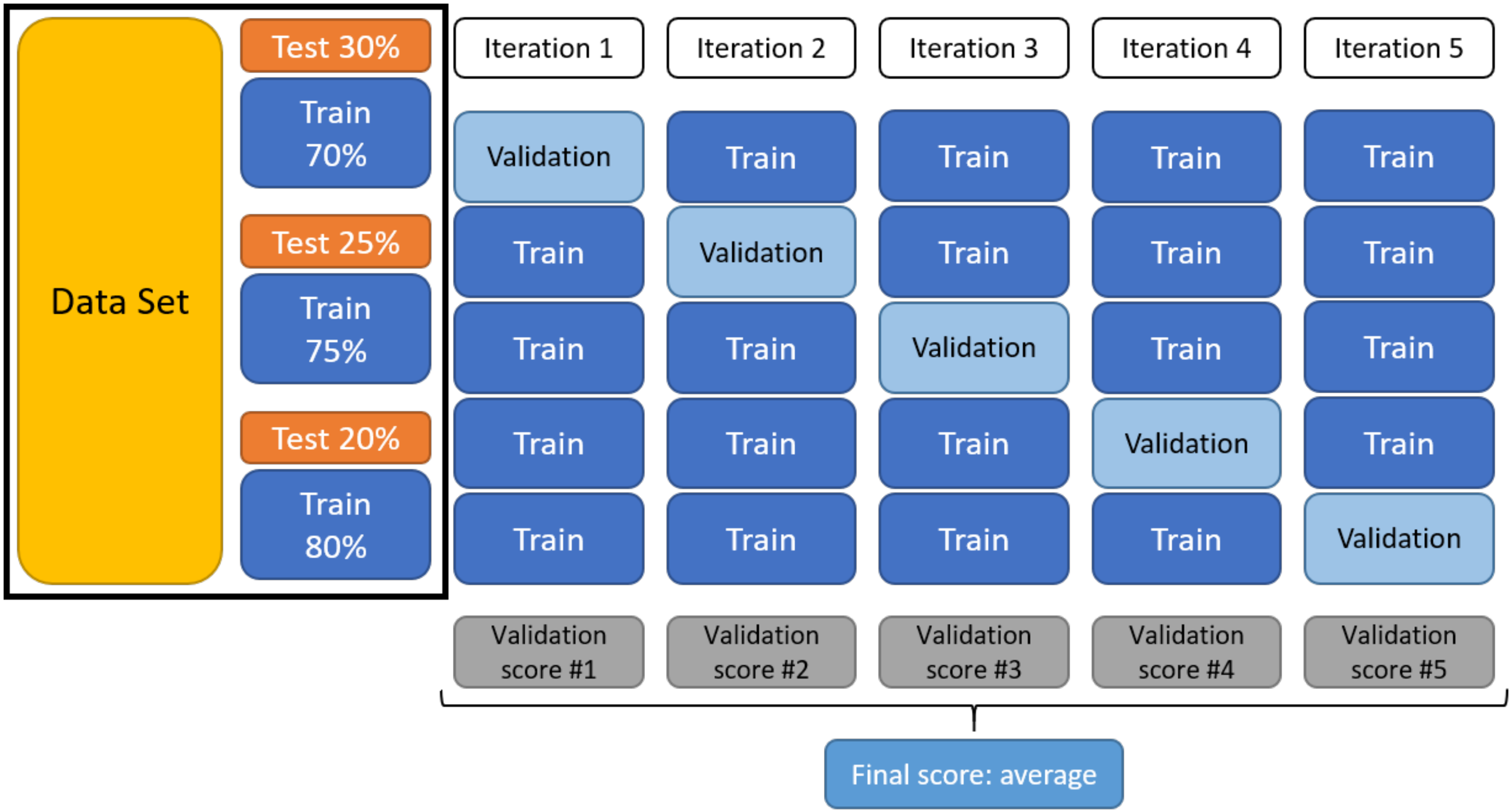


Figure 24: Cross Validation 5-Folds

Classification Algorithms (1/3)
(Support Vector Machine (SVM))

Minimize $\frac{1}{2}\ ||\vec{w}||^2 + C\sum_{i=1}^{N}\xi_i$ subject to $y_i(\vec{w} * \vec{x} + b) \geq 1 - \xi_1$ for i=1,…,N

$$Gamma = K\left(x_i, x_j\right) = \exp(-\gamma||x_i\text{-}x_j||^2),\ \gamma > 0$$

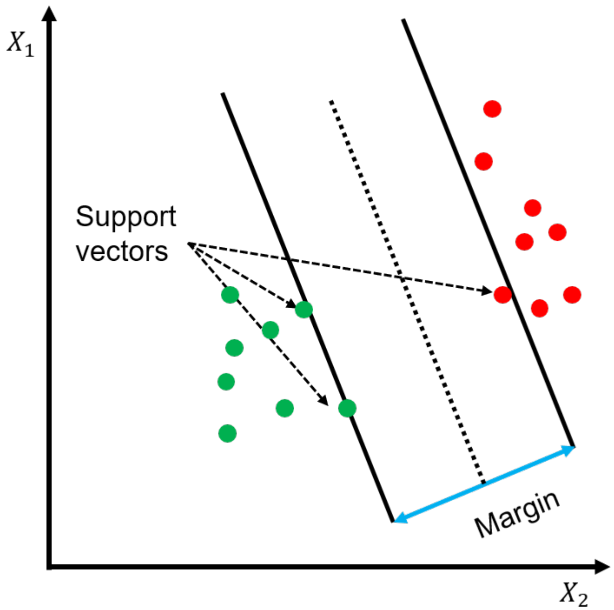


Figure 25: SVM

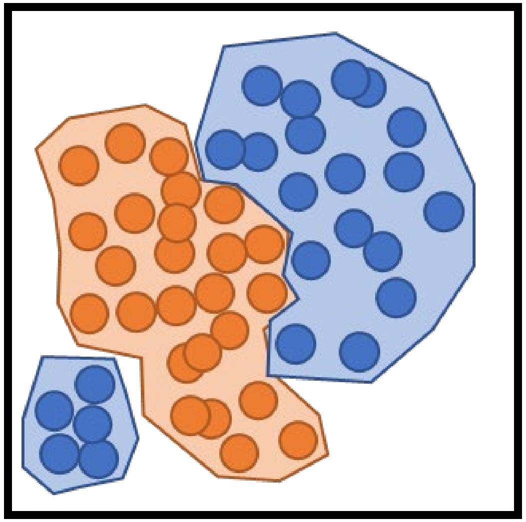


Figure 26: RBF Kernel

Classification Algorithms (2/3)

(Neural Networks (NN))

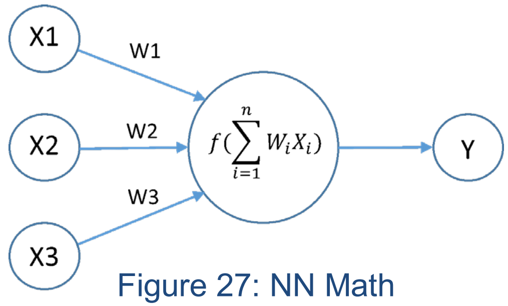


Figure 27: NN Math

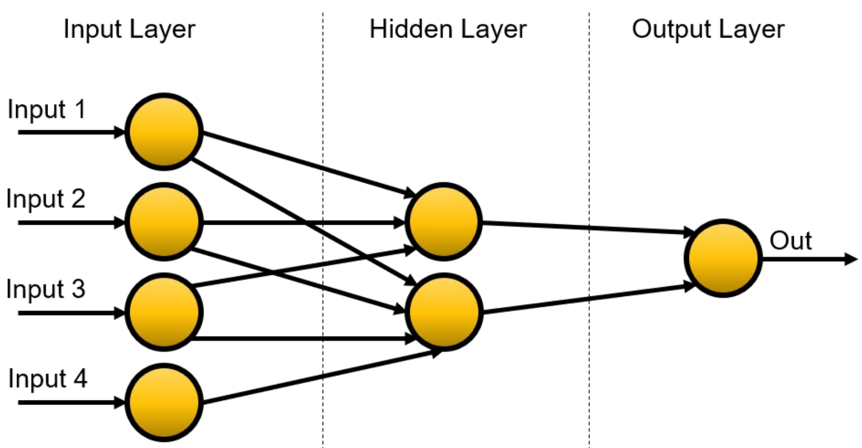


Figure 28: Neural Network

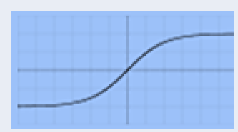
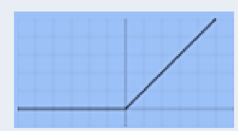
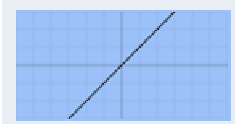
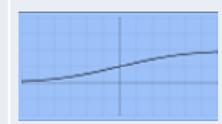
Hyperbolic Tangent	ReLu	Identity	Logistic
$\frac{2}{1+e^{-2x}} - 1$	$\begin{cases} 0 \; for \; x < 0 \\ x \; for \; x \geq 0 \end{cases}$	x	$\frac{1}{1+e^{-x}}$
			

Figure 29: NN Math

Classification Algorithms (3/3)

(K-Nearest Neighbor (KNN))

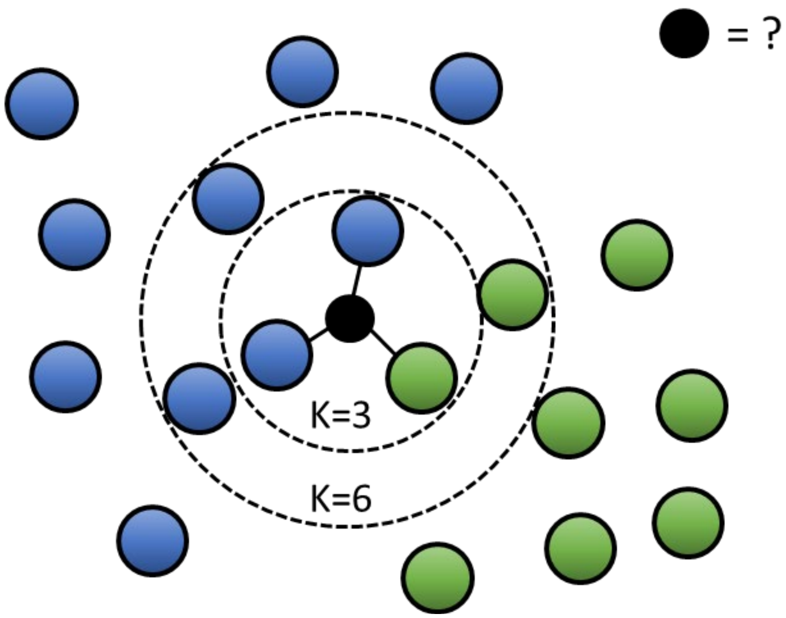


Figure 30: KNN

Distance Function

$$Euclidean = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Here:

n = n° of dimensions (2 in this case)
x = datapoint from dataset
y = new data point (to be predicted)

Performance Measures

Predicted Class \ Actual Class	Positive	Negative
Positive	True positive (TP)	False positive (FP)
Negative	False Negative (FN)	True Negative (TN)

Figure 31: Confusion matrix

TP: It's the class the models says

TN: It's the class the models says

FN: It's the class the models says

FP: It's the class the models says

Precision

$$p = \frac{TP}{TP + FP}$$

Recall

$$r = \frac{TP}{TP + FN}$$

F_1score

$$F_1\text{score} = 2\,\frac{p\,.\,r}{p + r}$$

Results (1/4)

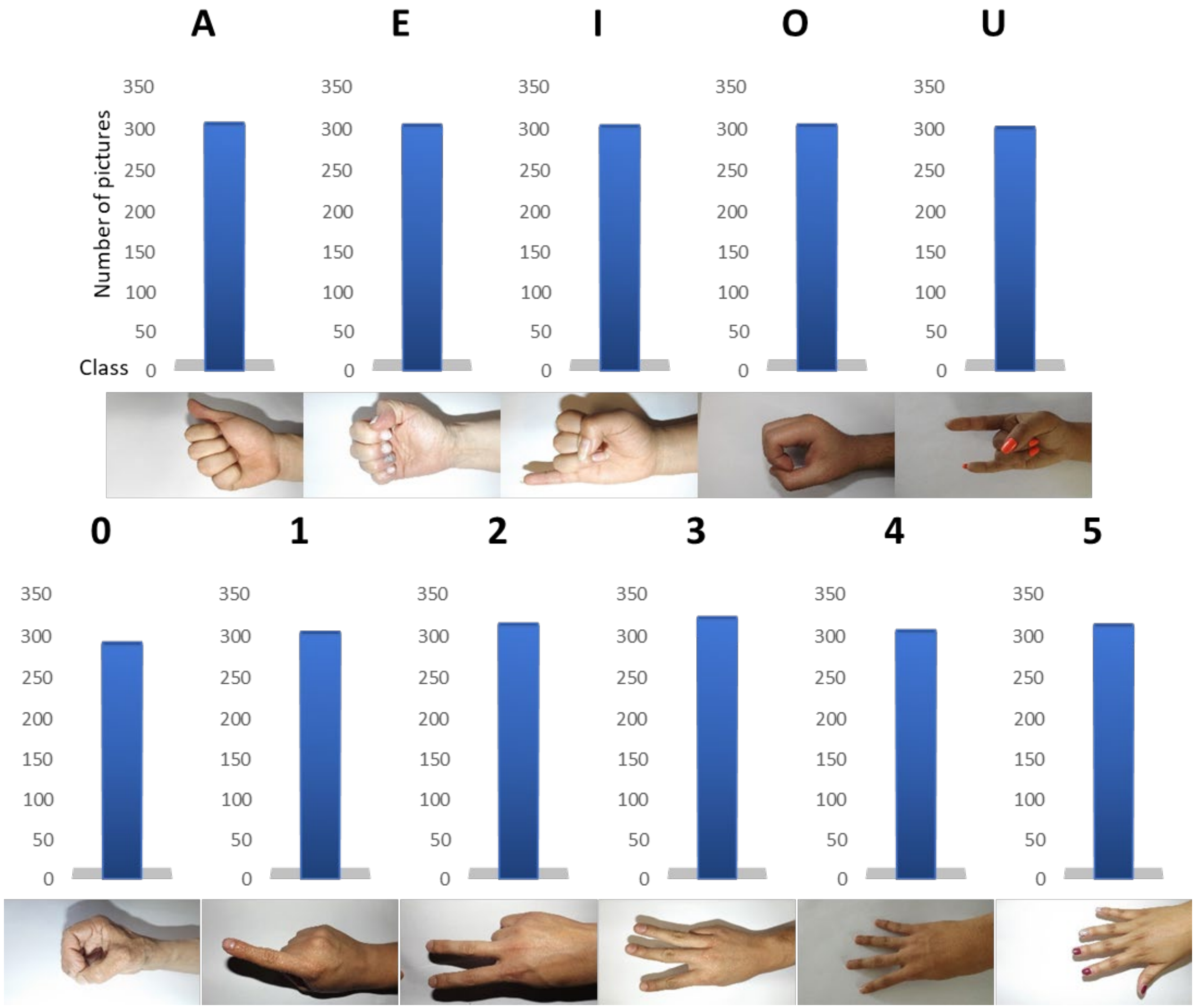


Figure 32: Dataset

Results (2/4)

P:Precision
R:Recall
F1:F1-Score

%Train - %Test	Method	SVM			KNN			NN		
		P	R	F1	P	R	F1	P	R	F1
70% - 30%	EF	0.67	0.67	0.67	0.62	0.61	0.61	0.41	0.37	0.35
	HOG	0.64	0.64	0.64	0.52	0.47	0.47	0.53	0.53	0.53
	HOG-PCA	0.64	0.63	0.63	0.52	0.47	0.46	0.49	0.47	0.47
	HU	0.28	0.25	0.21	0.30	0.30	0.29	0.21	0.25	0.22
	GEOMETRIC	0.34	0.13	0.12	0.23	0.22	0.22	0.01	0.08	0.01
	COMBINED	0.48	0.13	0.11	0.23	0.23	0.23	0.03	0.11	0.05
	HOG-EF	0.69	0.68	0.68	0.62	0.61	0.62	0.39	0.38	0.37
	HOG-PCA_EF	0.69	0.69	0.69	0.62	0.61	0.62	0.43	0.44	0.42
75% - 25%	EF	0.67	0.66	0.66	0.63	0.61	0.62	0.35	0.36	0.33
	HOG	0.66	0.65	0.65	0.54	0.48	0.48	0.54	0.53	0.52
	HOG-PCA	0.65	0.64	0.65	0.54	0.50	0.49	0.51	0.49	0.49
	HU	0.31	0.26	0.23	0.28	0.28	0.28	0.22	0.23	0.20
	GEOMETRIC	0.30	0.13	0.12	0.22	0.23	0.22	0.06	0.10	0.03
	COMBINED	0.41	0.14	0.11	0.23	0.23	0.23	0.01	0.09	0.02
	HOG-EF	0.68	0.67	0.67	0.63	0.62	0.62	0.43	0.43	0.41
	HOG-PCA_EF	0.69	0.67	0.68	0.63	0.62	0.62	0.46	0.48	0.46
80% - 20%	EF	0.68	0.67	0.68	0.64	0.62	0.63	0.28	0.32	0.28
	HOG	0.66	0.66	0.66	0.53	0.46	0.46	0.54	0.51	0.52
	HOG-PCA	0.66	0.65	0.65	0.54	0.47	0.48	0.47	0.46	0.46
	HU	0.27	0.25	0.22	0.29	0.28	0.28	0.23	0.24	0.23
	GEOMETRIC	0.33	0.12	0.13	0.22	0.22	0.22	0.05	0.17	0.07
	COMBINED	0.47	0.12	0.11	0.23	0.23	0.22	0.01	0.09	0.02
	HOG-EF	0.70	0.69	0.69	0.64	0.63	0.63	0.24	0.27	0.25
	HOG-PCA_EF	0.69	0.68	0.69	0.64	0.63	0.63	0.43	0.45	0.42

Figure 33: All result of performance

EF: Elliptic Fourier
HOG: Histogram of Oriented Gradients
HOG_PCA: Histogram of Oriented Gradients with Principal Component Analysis
COMBINED: Concatenate 4 Features

Result (3/4)
The best result

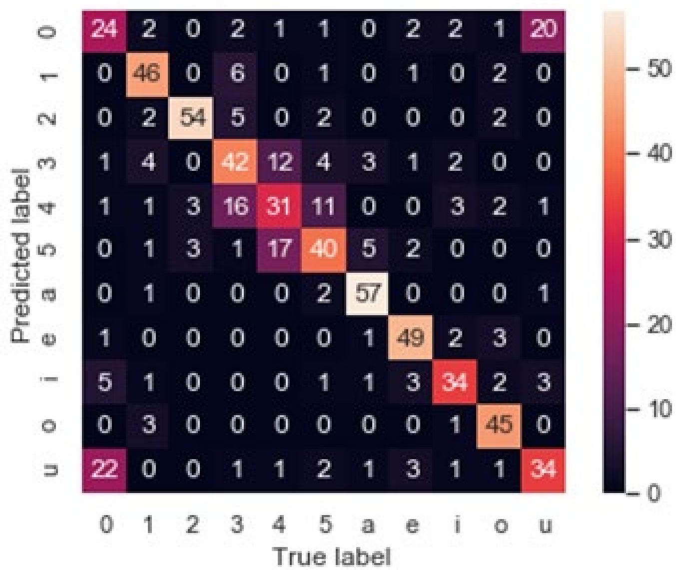


Figure 34: Confusion Matrix HOG and Elliptic Fourier in SVM

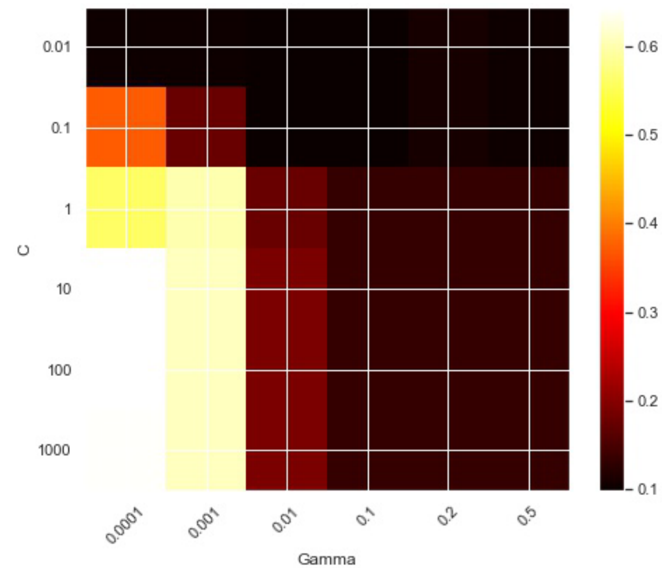


Figure 35: Heat map HOG and Elliptic Fourier in SVM

%Train - %Test	Method	SVM		
		P	R	F1
80% - 20%	HOG-EF	0.70	0.69	0.69

Figure 36: result of the best classifier

Results (4/4)

Classification Strategy for LSC's Hand Recognition Process

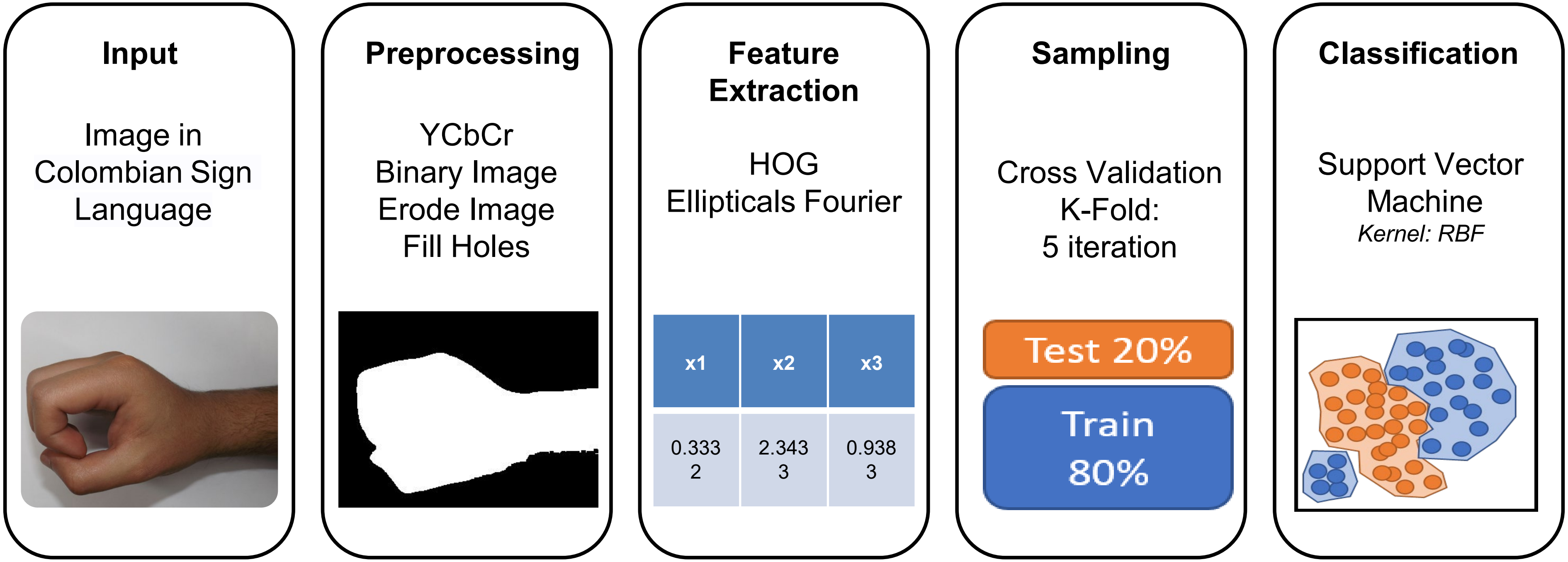


Figure 37: Classification Strategy

DISCUSSION

- According to the results table we can see that the best method for ***Support Vector Machines*** are **Histogram of Oriented Gradients** with **Fourier Ellipticals** with a percentage of 70%, 69%, 69% accuracy, recall and F1-score respectively using the 80% training set and 20% test set.

- It is discussed the margin of error that is presented where the confusion between the vowel **U** and the number **0**.

- As the **Support Vector Machine** were the best method of classification, perhaps by varying the parameters of ***Gamma*** and ***C***, you can get a better measure of performance.

CONCLUSION

- This experiment discusses several methods for the Recognition of Hand Gestures, including *Support Vector Machine*, *Neural Networks* and the *K-Nearest Neighbor*. Being the most precise ***Support Vector Machine*** when it comes to classifying or recognizing gestures.

- In the process of feature extraction, the features (HOG) were the most relevant as detailed information of each one of the images.

- The Morphological features did not give very good result because most of the images have similar characteristics like the area, the contour, among others. For this reason the algorithm confused or predicted the signals in a bad way.

- In the cross-validation process, after testing with 70%-30%, 75%-25% and 80%-20% for training-testing respectively. The best data partition was 80% for training and 20% for data testing.

PRODUCTS (1/2)

- **Dataset**: The first dataset of Colombian sign Language data (LSC) is delivered where the vowels and numbers from 0 to 5 are found. 3324 photos were taken for the construction of the data set (**11GB)** : **https://goo.gl/9uSMm1**

- **Repository**: A repository is delivered where all the experiment is carried out for didactic, research and other purposes. Using the MIT (Massachusetts Institute of Technology) license. Leaving everything OPEN SOURCE mode:**https://goo.gl/qP39SN**

- **Paper:** A scientific article (Paper) is given, in which the experimental process carried out in the thesis is described: **https://goo.gl/xcMSaQ**

Figure 38: Icons

PRODUCTS (2/2)

- **App Mobile (Android):** A mobile application is delivered, made for the Android operating system. Where the experiment is carried out from the app. Using two scientific methods of machine learning, such as TensorFlow and Google Cloud Vision: https://goo.gl/jSb43X

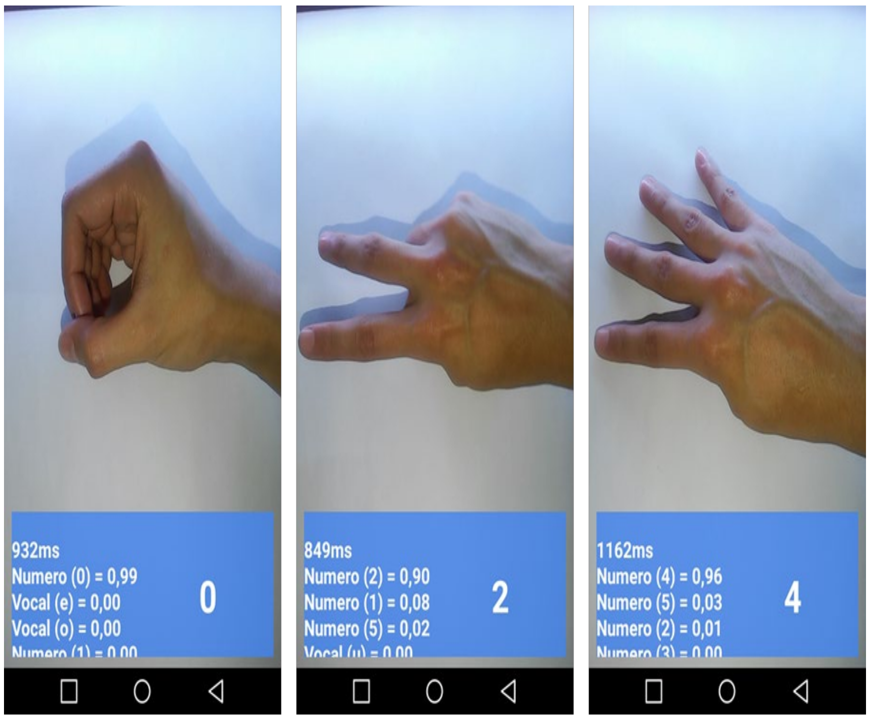


Figure 39: Recognizing the numbers

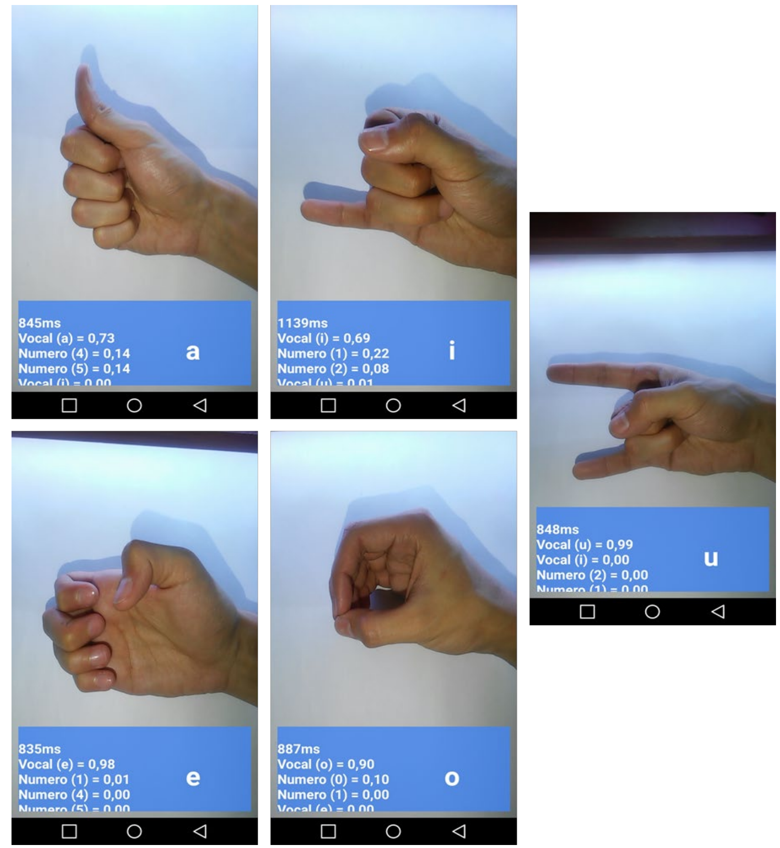


Figure 40: Recognizing the vowels

FUTURE WORKS

- It is recommended to perform the experiment using other classification techniques such as neural networks convolutional, Naive Bayes, Logistic regression, among others. In order to buy performance and accuracy of predictions, as you can get better results with other classification methods.

- It is recommended to increase the data set so that the algorithm improves your learning and training process.

- It is recommended to expand the dataset to other cities in the world to be the most comprehensive dataset. Taking into account the skin color of all people.

- It is recommended to carry out the development of the mobile application for iOS OS and upload it to the Apple store called App Store so that the experiment can be tested from both OS (Android and iOS)

- A great future job would be to perform the process of recognizing videos for Colombian sign language.

BIBLIOGRAPHY

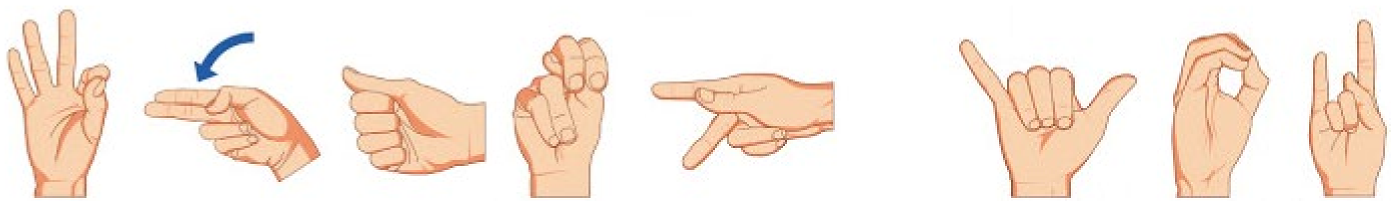
- Douglas, C., & King N, N. (1999). *Face Segmentation Using Skin-Color Map in Videophone Applications* [Ebook] (p. 5). Retrieved from https://pdfs.semanticscholar.org/8055/dcb4e491dfe35b72e7817ce51e12184ce608.pdf

- ABC. (2013). Un sordomudo mata a su mujer tras avisar a su hermano por SMS de que iba a hacerlo. Retrieved from https://www.abc.es/sociedad/20130120/rc-sordomudo-mata-mujer-tras-201301201110.html

- 9GAG. (2018). *Y es por eso que ver un problema desde diferentes ángulos es tan importante*[Red social]. Retrieved from https://m.9gag.com/gag/anMOvy5

- Fundación Internacional, M. (2015). *Alfabeto Dactilológico Colombiano* [Video]. Fundación Internacional María Luisa de Moreno • FIMLM. Retrieved from https://www.youtube.com/watch?v=X3UuOWOaulA

- Ayudas Para Todos. (2013). *Vocabulario Básico 1 - Abecedario Lengua de Señas Colombiana LSC* [Video]. Retrieved from https://www.youtube.com/watch?v=JMraBJsA9oI



T H A N K   Y O U

UNIVERSIDAD CATÓLICA de Colombia
Vigilada Mineducación