الجمهورية الجزائرية الديمقراطية الشعبية
Peoples Democratic Republic of Algeria
وزارة التعليم العالي و البحث العلمي
Ministry of Higher Education and scientific research
المدرسة الوطنية العليا للتكنولوجيات المتقدمة
National High School of Advanced Technology
Department of Electrical Engineering and Industrial Computer Science

Final Year Project to Obtain the Diploma of Engineering (Master)

Field:

Electronics

Specialty:

Embedded Systems

Subject:

Comparative Analysis of Random Forest, YOLO, and Classical Lookup Table-based Approaches for Sign Language Classification

Realized by

Benbrahim Tarek

Examination Committee:

Chair	Zellat Khadidja	ENSTA
Supervisor	Bougherira Hamida	USD Blida
Co-Supervisor	Habani Lamia	ENSTA
Examiner	Bouchama Samira	ENSTA
Examiner	Kheira Lakhdari	ENSTA

Algiers, The 26/06/2024

Academic year 2023 - 2024

Comparative Analysis of Random Forest, YOLO, and Classical Lookup Table-based Approaches for Sign Language Classification

Tarek Benbrahim[1] Bougherira Hamida[2]
Habani Lamia[3]
National Higher School of Advanced Technologies (ENSTA)
Algiers, Algeria

Abstract

This article presents a comparative analysis of three sign language classification methods: random forest, YOLO (You Look Once), and look-up table method , the strengths and weaknesses of each method are examined diversity in terms of accuracy, speed, flexibility.

1 Introduction

Sign language classification plays an important role in bridging the communication gap between the deaf and hard of hearing. As technology advances, techniques have been developed to accurately translate sign language gestures into readable text or spoken words This article examines and compares three distinct ways of classifying sign language: random forest, YOLO (You Only Look Once), and the usual Lookup Table -based method

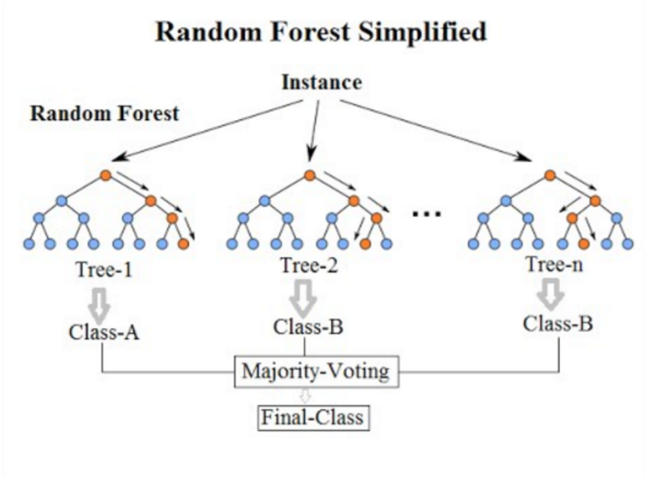


Figure 1: Random Forest architecture diagram

2 Random Forest

Random Forest is a supervised machine learning algorithm which consists of an ensemble of Decision Trees whereby the final/leaf node will be either the majority class for classification problems or the average for regression problems. A random forest will grow many classification trees and for each output from that tree, we say the tree 'votes' for that class. [1]

3 YOLO

YOLO is a deep learning based architecture.It is a real-time object recognition system that processes images using a single neuron.The YOLO system is based on convolutional neural networks (CNNs) that enable high speed detection and high accuracy as it divides the image into zones and determines the bounding boxes and probabilities for each zone. [2]

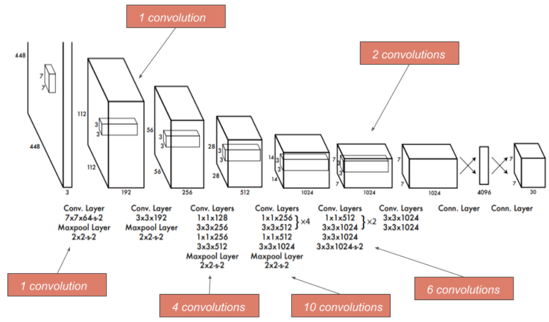


Figure 2: Yolo architecture diagram

4 Lookup tables

A lookup table (LUT) is an array of data that maps input values to output values, thereby approximating a mathematical function. Given a set of input values, a lookup operation retrieves the corresponding output values from the table.[3]

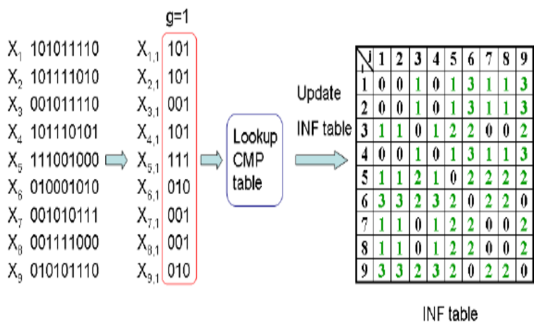


Figure 3: Lookup Tables architecture diagram

5 Methodology

5.1 Random Forest

First of all , we made a smart glove that consists of five flex senosrs , We trained two random forest models for ASL (American sign language) classification , one to predict sign language alphabet whereas the other one is to predict sign language expressions. The number of features is 11 : five for the flex sensors and 6 derived from the MPU6050 (Accelerometer & Gyroscope).

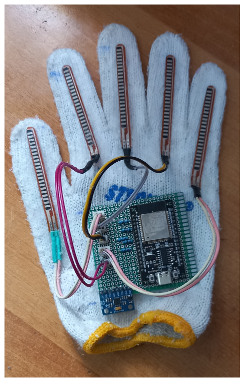


Figure 4: Smart glove

For the first one , we got an accuracy of 82% and only 72% for the second one , the Figure 4 and Figure 5 show the changing of accuracy according to the number of estimators (number of trees) for both models

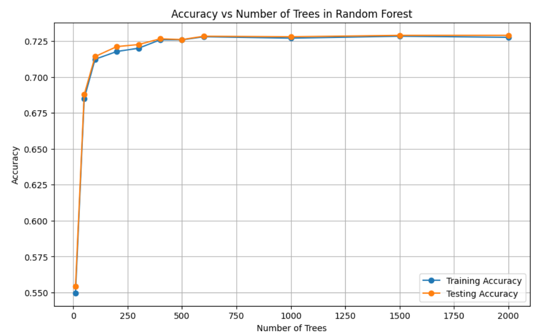


Figure 5: Training and testing of sign language expressions model

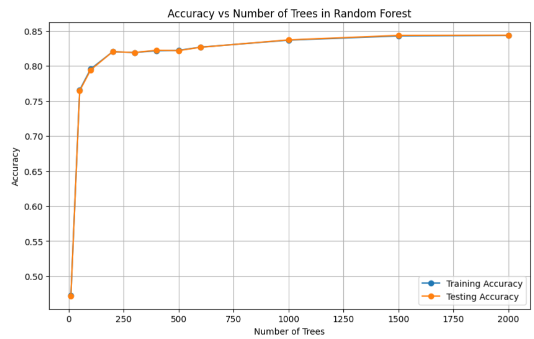


Figure 6: Sign language alphabet training and testing model

The 1 shows a sample of the dataset used :

flex_1	flex_2	flex_3	flex_4	flex_5	GYRx	GYRy	GYRz	ACCx	ACCy	ACCz
0	46	47	40	37	-0.022901	-0.015267	-0.015267	9.380102	2.204761	-0.616089
0	47	44	38	35	-0.022901	-0.015267	-0.015267	9.347803	2.209546	-0.618481
0	47	48	46	33	-0.015267	-0.015267	-0.007634	9.317896	2.183228	-0.634033
0	48	50	40	33	-0.015267	-0.007634	-0.007634	9.326269	2.154517	-0.662744
0	47	47	41	36	-0.007634	-0.007634	-0.007634	9.34541	2.18562	-0.679492
0	45	50	44	36	-0.007634	-0.007634	0.0	9.348999	2.228687	-0.697437
0	49	46	43	37	0.0	-0.007634	0.0	9.353785	2.263379	-0.709399
0	51	48	42	38	-0.007634	-0.007634	0.007634	9.363355	2.321997	-0.704614
0	48	48	38	35	-0.015267	-0.015267	0.015267	9.369336	2.42727	-0.680689
0	52	44	40	37	-0.022901	-0.015267	0.022901	9.374121	2.534936	-0.655566
0	49	48	42	37	-0.022901	-0.022901	0.022901	9.386084	2.612695	-0.620874
0	49	48	40	35	-0.022901	-0.022901	0.015267	9.392065	2.647388	-0.588574
0	48	46	38	33	-0.015267	-0.022901	0.007634	9.414795	2.639014	-0.551489
0	49	47	41	35	-0.015267	-0.015267	0.0	9.473413	2.604321	-0.508423
0	47	47	40	35	-0.007634	-0.015267	-0.007634	9.527246	2.558862	-0.446216
0	48	47	40	35	-0.015267	-0.022901	-0.015267	9.563135	2.511011	-0.385205

Table 1: Dataset Sample

We can see the that the accuracy for both models increase each time the number of trees increase as well and then it stabilizes,the more the trees are , the more number of "voters" are as each tree is trained separately and then the decision is made by aggregating all the votes of the estimators which explains the accuracy's enhancement.

The Table 5.1 shows the classification report of all the classes (alphabet).

The support number is almost the same for all the classes due to the equal number of entries of the dataset given to all labels.

As we see, Most of them are well classified , we can spot some classes that have poor classification like the class "S" (0.67) , The class "E" (0.71) , the reason is for example these two mentioned classes have slightly the same sign language gestures thus they are correlated between each other thus the machine learning model finds some difficulties distinguishing between them. in addition , we plotted the correlation matrix for both classes as seen in the **Figure 6** and **Figure 7**:

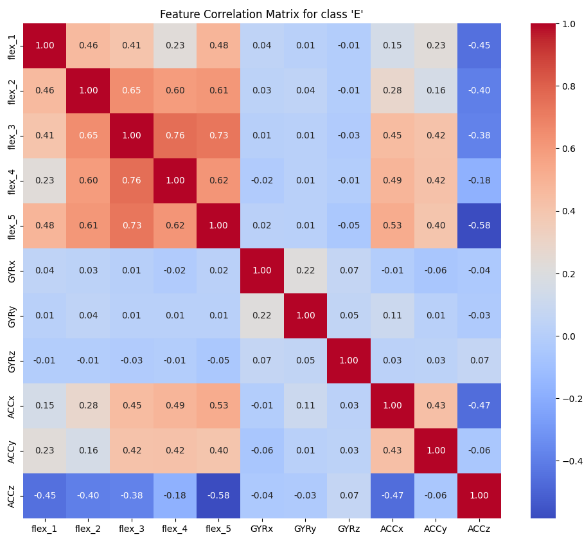


Figure 7: correlation plot E

The correlation coefficient in class 'E' between 'flex 1' and 'flex 2' is 0.46, suggesting a moderately positive linear association. This indicates that 'flex 2' tends to increase in value as 'flex 1' increases.

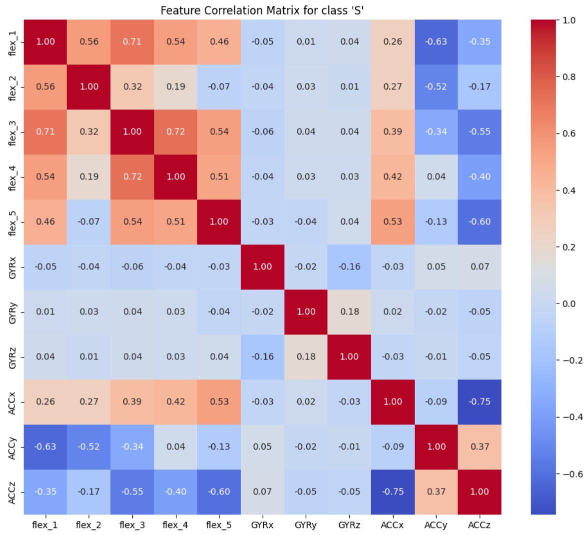


Figure 8: correlation plot S

The correlation coefficient in class 'S' between 'flex 1' and 'flex 2' is 0.56, suggesting a stronger relationship compared to class 'E'. This indicates that in the sign language gesture for the letter 'S', the values of 'flex 1' and 'flex 2' are more closely related or influence each other more significantly than they do in the gesture for the letter 'E'.

In addition to that , the two correlation matrices are mostly the same which explains the model's difficulty distinguishing between these correlated classes.

	precision	recall	f1-score	support
A	0.74	0.89	0.81	3204.0
B	0.99	0.98	0.99	3155.0
C	0.82	0.70	0.75	3227.0
D	0.66	0.97	0.78	3170.0
E	0.63	0.83	0.71	3166.0
F	0.99	1.00	0.99	3122.0
G	0.86	1.00	0.93	3102.0
H	1.00	1.00	1.00	3163.0
I	0.76	0.99	0.86	3068.0
J	1.00	0.67	0.80	3183.0
K	0.73	0.79	0.76	3174.0
L	0.83	0.98	0.90	3072.0
M	0.87	0.93	0.90	3141.0
N	0.86	0.82	0.84	3013.0
O	0.79	0.74	0.76	3182.0
P	0.99	1.00	0.99	3172.0
Q	0.85	1.00	0.92	3176.0
R	0.86	0.72	0.78	3164.0
S	0.84	0.56	0.67	3198.0
T	0.96	0.57	0.71	3140.0
U	0.75	0.86	0.80	3161.0
V	0.81	0.79	0.80	3234.0
W	0.98	1.00	0.99	3157.0
X	0.87	0.76	0.81	3070.0
Y	0.92	1.00	0.96	3142.0
Z	0.99	0.43	0.60	3144.0
accuracy	0.84	0.84	0.84	0.84
macro avg	0.86	0.84	0.83	81900.0
weighted avg	0.86	0.84	0.84	81900.0

Table 2: sign language alphabet Classification report

5.2 YOLO

Shobhit Tyagi ,Prashant Upadhyay ,Hoor Fatima , Sachin Jain , Avinash Kumar Sharma trained a yolov5 and yolov8 models for american sign language classification , for the first model , the mean average precision (mAP) for each model was 93.6& and 96%. The dataset used contains images for testing, training and validation in the ratio of 1: 21: 2 respectively (72 for testing, 1512 for training, 144 for validation) and the new model has been trained for 80 epochs. The **Figure 8** and **Figure 9** shows the mAP for The class "Q" and class "A".[4]

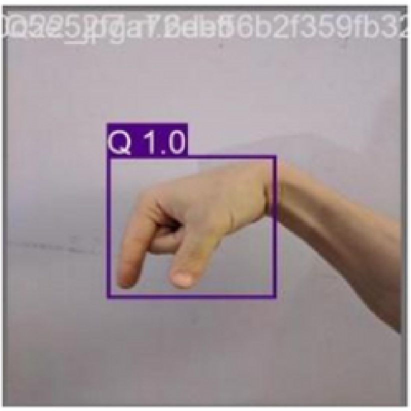


Figure 9: Letter Q in sign language [4]

Figure 10: Letter A in sign language [4]

The authors also mentioned the initial and final loss of the two models, expressing that in YOLOv8's case, those occurrences

Initial and Final loss

	BOUNDING BOX LOSS		CLASSIFICATION LOSS		mAP
	1st epoch	80th epoch	1st epoch	80th epoch	
YOLOv5	1.733	0.3265	4.251	0.191	93.6%
YOLOv8	1.444	0.312	4.352	0.1735	96%

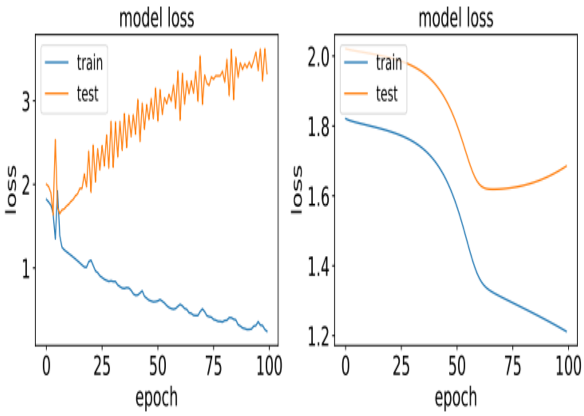
Figure 11: Initial and final loss [4]

are 13. While this comparison may seem redundant,this observation may be attributed to YOLOv8 being in its early stages of development compared to its predecessor. Also , both models could perform well when the two gestures are somehwat the same. [4]

5.3 Lookup tables

Jennifer J. Gago ,Valentina Vasco ,Bartek Łukawski ,Ugo Pattacini , Vadim Tikhanoff ,Juan G. Victores and Carlos Balaguer used the lookup-table algorithms to implement spanish sign language into a humanoid. The process starts first by training a RNN model that convert the written spanish to LSE tokens (Lengua de Señas Española) , and then using these tokens to create a lookup-table so the humanoid use this configuration to to perform the corresponding sign language gestures.[5]

The lookup table configuration is obtained from recording joints and limbs movement using motion capture technology and then these records are converted to produce 2D and 3D skeleton pose estimation.[5]



(a) Learning rate $\alpha = 0.0001$ (b) Learning rate $\alpha = 0.001$

Figure 12: learning rate [5]

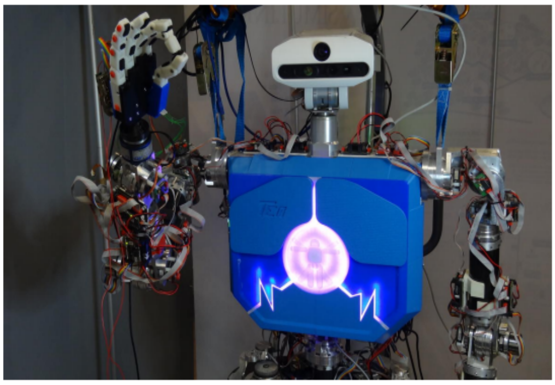


Figure 13: Teo the humanoid [5]

According to the authors , the RNN model showed high accuracy as it could predict most of the tokens.

6 Comparative analysis

We compared between the three algorithms , now , we can mention some advantages and disadvantages of each one :

1. Random Forest algorithms show high accuracy with tabular dataset , the ability to avoid overfitting and handle both regression and classification tasks which makes it suitable for many machine learning applications[6]. one of its disadvantages is its hyperparameters , as high accuracy requires many number of trees which affects the speed of training.[7]

2. Yolo showed better accuracy compared to random forest model in sign language classification as it could distinguish between alphabet that have somewhat the same gestures which Random forest model struggled to attain it , but it requires a large dataset compared to the first approach.[8]

3. Lookup-tables also showed impressive results when implemented with a deep learning model for sign language classification but they lack flexibility as they can't be used as the primary method for that type of classification , however , they don't require large dataset which can be beneficial sometimes when the focus is on memory's storage. [9]

7 Conclusion

In conclusion, the best approach for sign language classification depends on the specific requirements and constraints of the task. If robustness and high accuracy are required, a deep learning method like YOLO may be preferred though potentially higher implementation time and computational resources required. On the other hand, if a balance between accuracy and implementation time is needed, a machine learning method like Random Forest could be a suitable choice.If the dataset isn't large or the number of sign gestures isn't numerous , lookup-tables may be the best approach for that.

References

[1] Leo Breiman. "Random forests". In: *Machine learning* 45.1 (2001), pp. 5–32.

[2] Zoumana Kelta. *YOLO Object Detection Explained: A Beginner's Guide*. Sept. 2022. URL: https://www.datacamp.com/blog/yolo-object-detection-explained.

[3] *About Lookup Table Blocks - MATLAB & Simulink*. URL: https://www.mathworks.com/help/simulink/ug/about-lookup-table-blocks.html.

[4] Shobhit Tyagi et al. "American Sign Language Detection using YOLOv5 and YOLOv8". en. In: (July 2023). DOI: 10.21203/rs.3.rs-3126918/v1. URL: https://www.researchsquare.com/article/rs-3126918/v1.

[5] Jennifer J. Gago et al. "Sequence-to-Sequence Natural Language to Humanoid Robot Sign Language". en. In: *EUROSIM 2019 Abstract Volume*. arXiv:1907.04198 [cs]. 2019. DOI: 10.11128/arep.58. URL: http://arxiv.org/abs/1907.04198.

[6] James Montantes. *3 reasons to use random forest over a neural network–comparing machine learning versus deep...* Apr. 2024. URL: https://towardsdatascience.com/3-reasons-to-use-random-forest-over-a-neural-network-comparing-machine-learning-versus-deep-f9d65a154d89.

[7] Niklas Donges. *Random Forest: a Complete Guide for Machine Learning*. July 2021. URL: https://builtin.com/data-science/random-forest-algorithm.

[8] Muhammad Asad Nadeem. *Unraveling YOLO v8: Benefits and Challenges in Object Detection*. Jan. 2024. URL: https://medium.com/@masadnadeem23/unraveling-yolo-v8-benefits-and-challenges-in-object-detection-1ec762debefd.

[9] *Understanding the Basics of Lookup Table Database in Database Management*. URL: https://kyligence.io/plp/understanding-basics-lookup-table-database-management/.