



ENSIM
École d'ingénieurs
Le Mans Université

Présentation du projet 5A :

TinyML pour l'IA embarquée

Réalisé par :

NAIT HAMMOU Mehdi
BEN BRAHIM Abdelkarim

Encadré par :

Pr. HASSAN Kais

+

Année Universitaire : 2021/2022

+ Sommaire

01

**Présentation
du projet**

03

**Etapes de
réalisation**

02

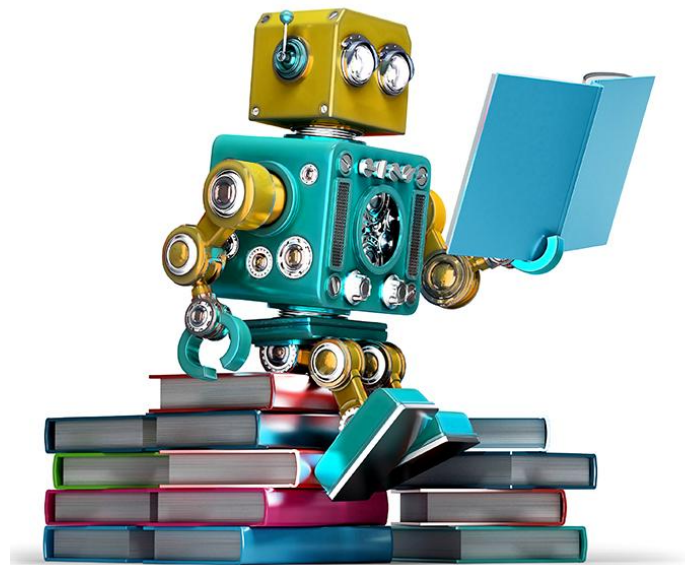
**Outils de
développement**

04

**Conclusion et
perspectives**

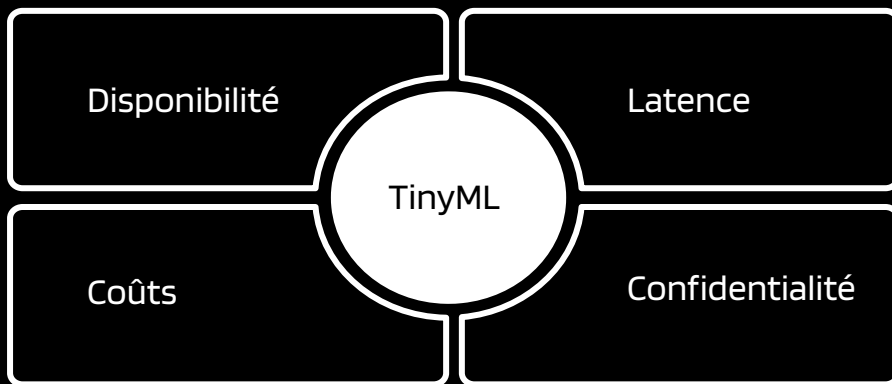
+ Introduction

TinyML, c'est quoi ?



Plusieurs problèmes, une solution

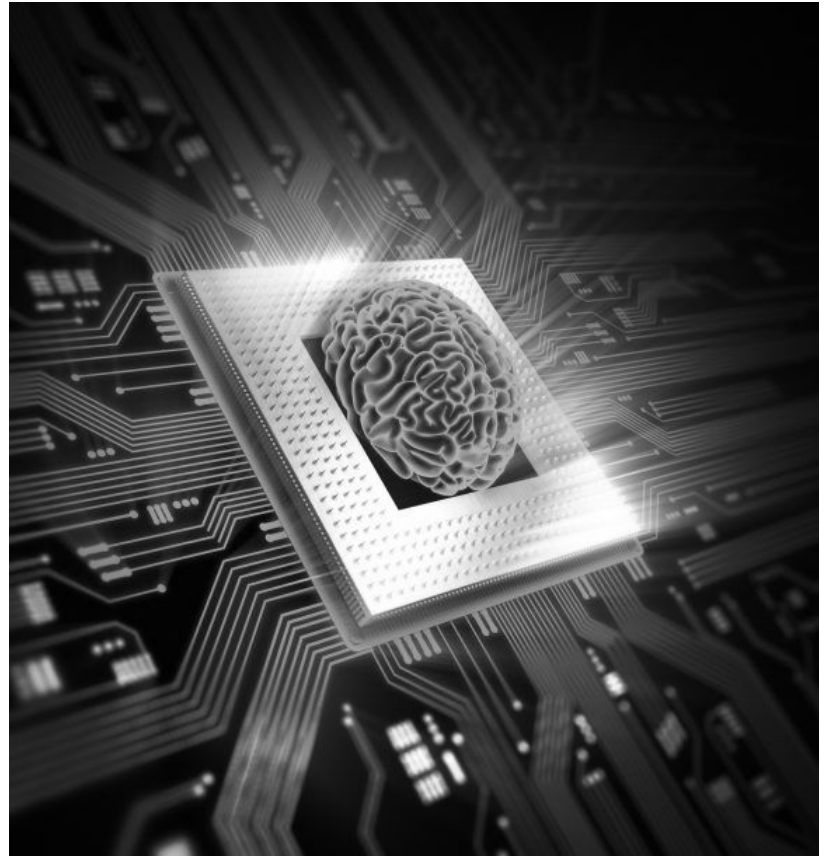
TinyML représente une solution pour diverses problèmes :



...

01

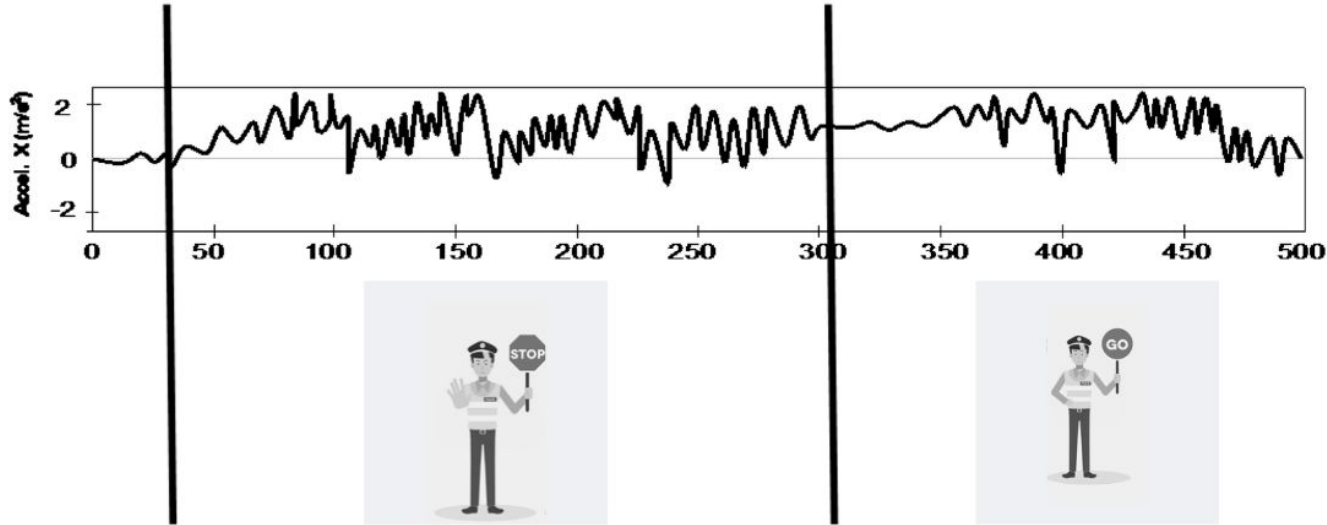
Présentation du projet



+ Présentation du projet



+ Présentation du projet



Le but de l'application est la reconnaissance des signes effectués par le policier.

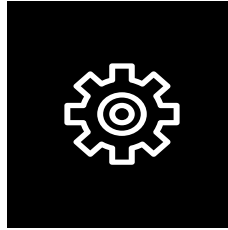
+ Objectifs du projet



Implémentation



Implémentation et
optimisation du modèle de
classification



Déploiement



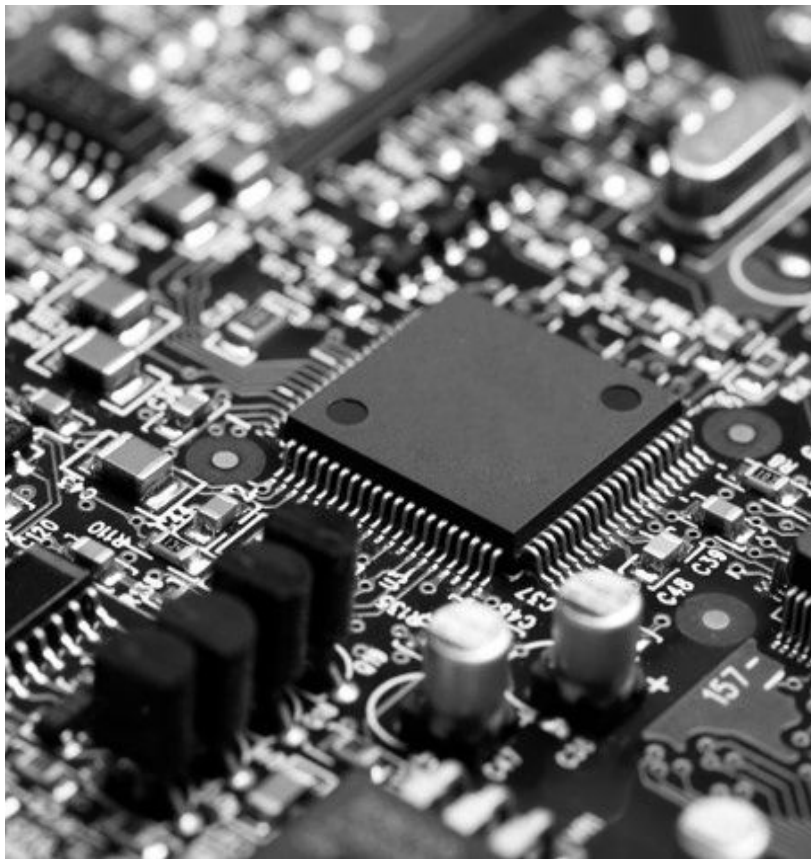
Déploiement du modèle et
exploitation des résultats
sur les cartes cibles



Documentation



Réalisation d'une
documentation détaillée
du projet

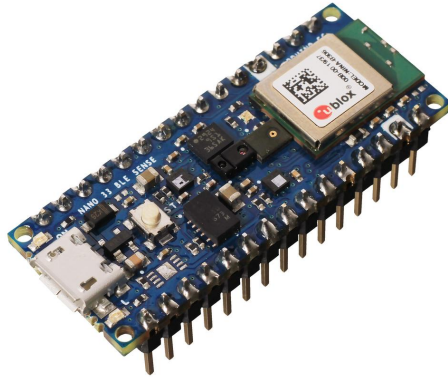


02

Outils de développement

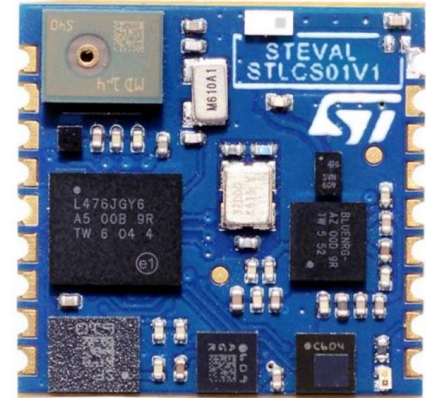
+ Outils de développement

Environnement Hardware :



Arduino Nano BLE 33 Sense

ARM Cortex-M4 64MHz
1MB de mémoire Flash
256KB de RAM
Floating point unit



STM SensorTile

ARM Cortex-M4 80 MHz
1MB de mémoire Flash
256KB de RAM
Floating point unit

+ Outils de développement

Environnement Logiciel :



TensorFlow Lite



PlatformIO



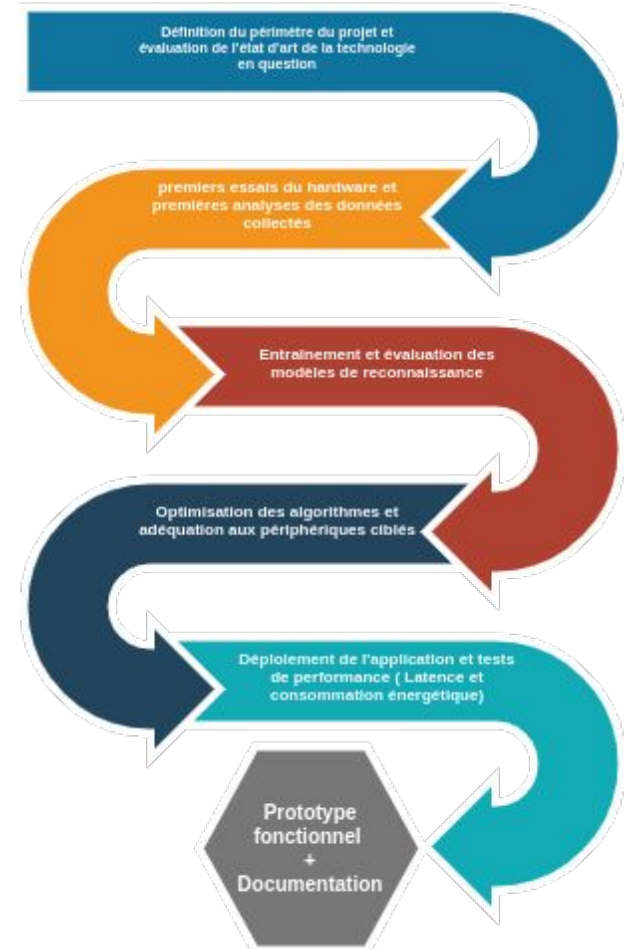
Tiny Motion Trainer

03

Étapes de réalisation

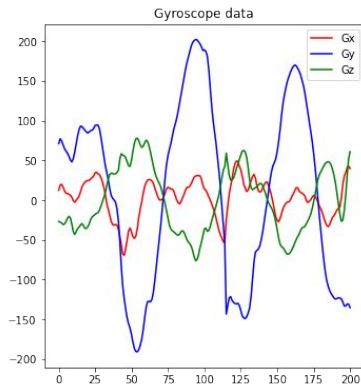
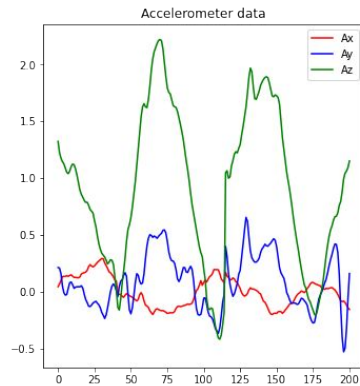
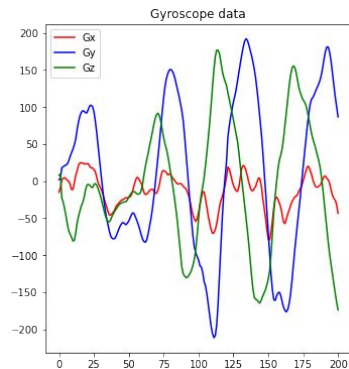
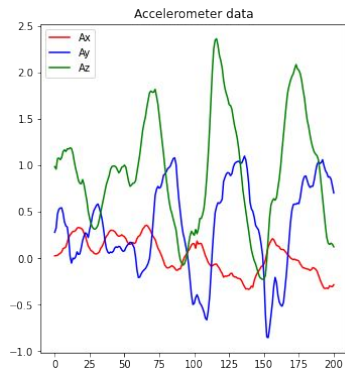


+ Etapes de réalisation

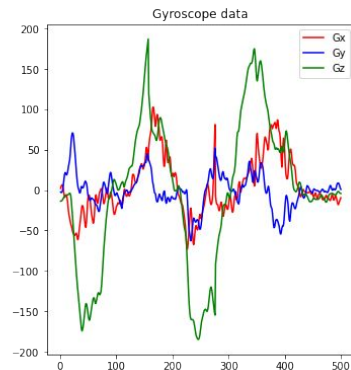
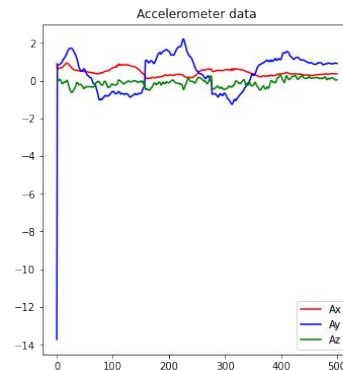


+ Capture des données

Signe d'accélération



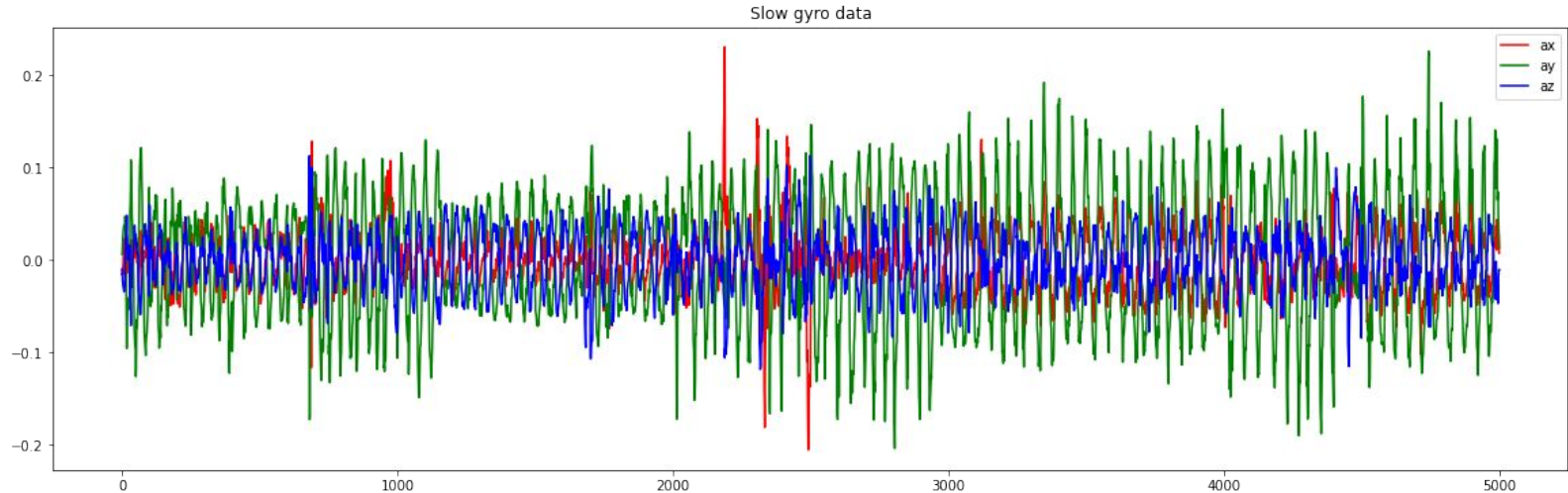
Signe d'arrêt



Signe de ralentissement

+ Capture des données

Construction de la dataset

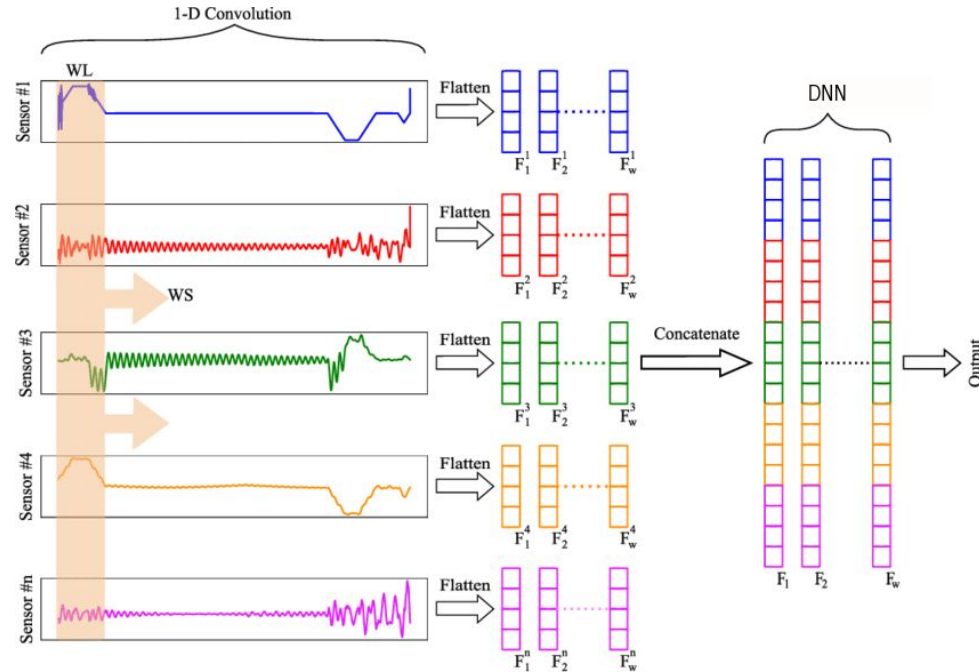


600*6 signal par mouvement

50 échantillon par signal, soit 0.5s de durée de signal

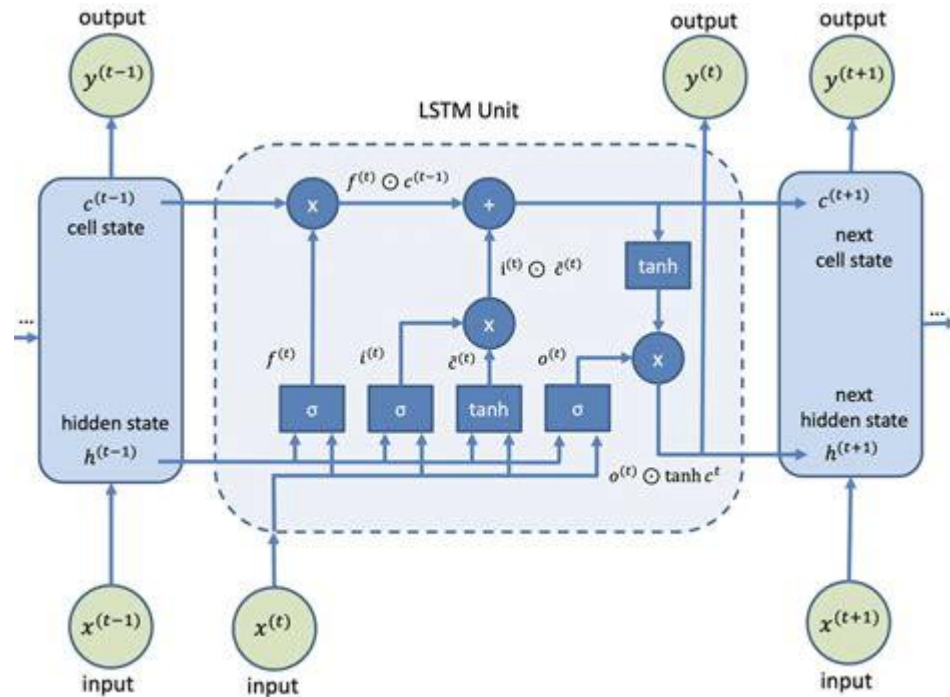
+ Architectures des réseaux de neurones

Réseau convolutif 1D



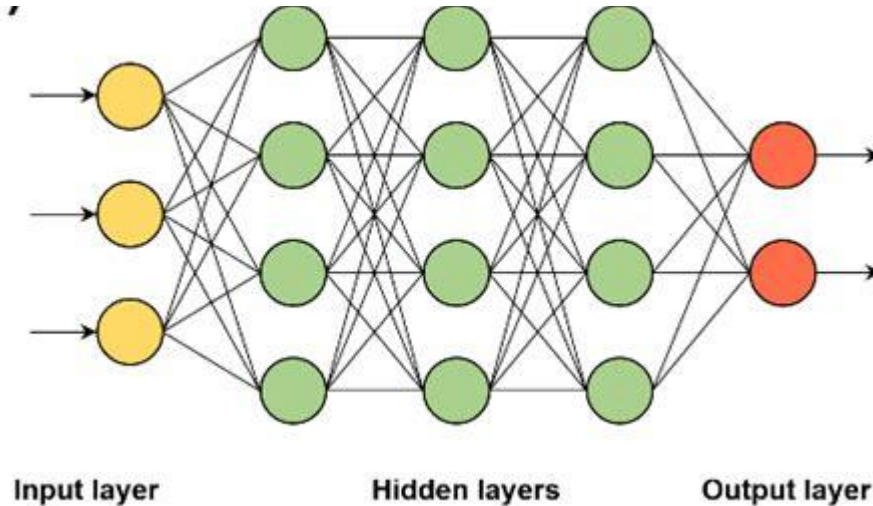
+ Architectures des réseaux de neurones

Long Short Term Memory (LSTM)



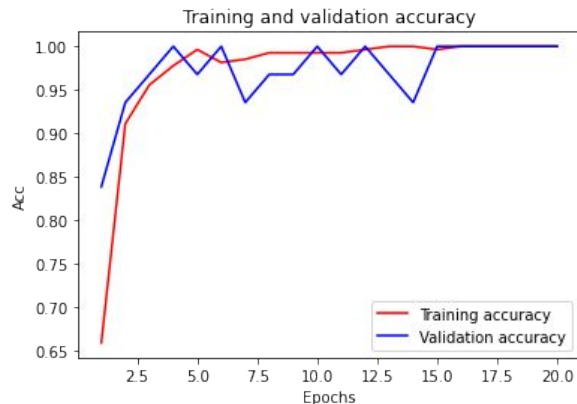
+ Architectures des réseaux de neurones

Baseline : Réseau de neurones dense standard

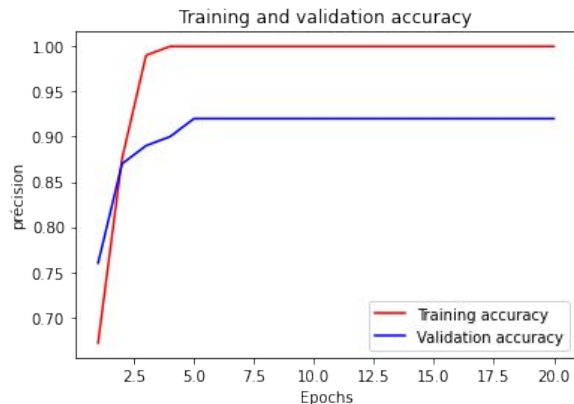


+ Résultats des entraînements

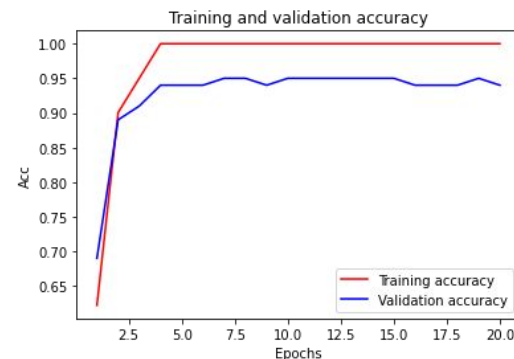
LSTM



DNN

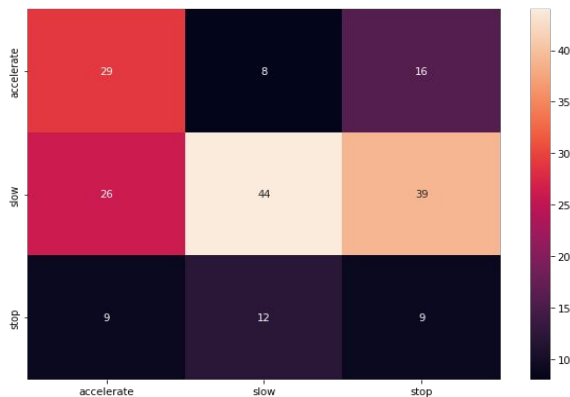


Réseau convolutif 1D

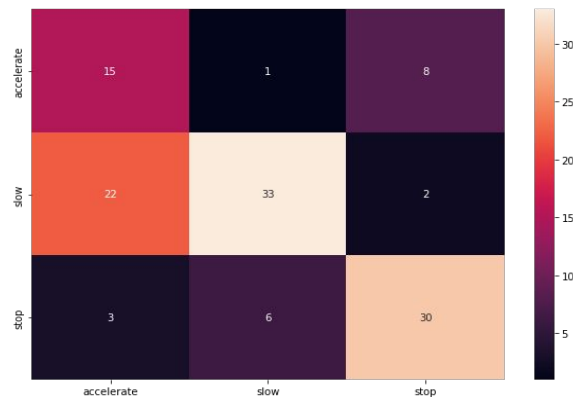


+ Validation des résultats

Réalité :



1D-CNN



LSTM

-> Grand problème de généralisation (Précision ~30%)

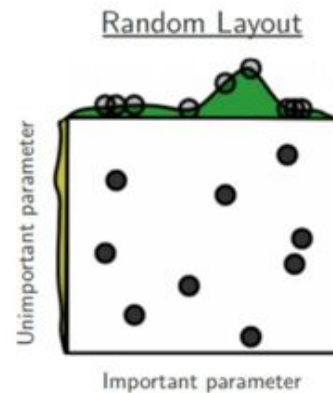
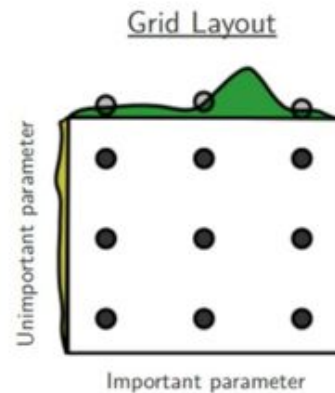
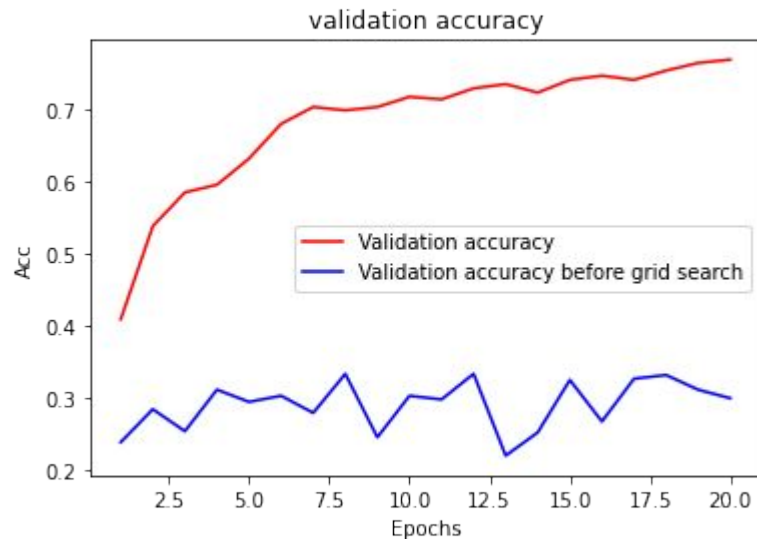
+ Validation des résultats

Solution : Revoir la stratégie de préparation de données

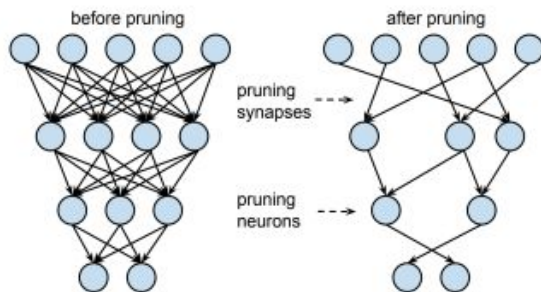


+ Validation des résultats

Ré-entraînement et optimisation des hyper-paramètres



+ Optimisation du modèle



Pruning

(a) Weights
after pruning

| | | | |
|-------|-------|-------|-------|
| -1.01 | 1.00 | 0.00 | 0.88 |
| 0.00 | 0.17 | 0.00 | -0.02 |
| 0.56 | 0.00 | 0.38 | 0.00 |
| 0.00 | -0.49 | -0.95 | 0.00 |

(b) Weights in
quantization levels

| | | | |
|-------|-------|-------|-------|
| -1.00 | 1.00 | 0.00 | 1.00 |
| 0.00 | 0.50 | 0.00 | -0.50 |
| 0.50 | 0.00 | 0.50 | 0.00 |
| 0.00 | -0.50 | -1.00 | 0.00 |

(c) Weights stored
in hardware

| | | | |
|----|----|----|----|
| -2 | 2 | 0 | 2 |
| 0 | 1 | 0 | -1 |
| 1 | 0 | 1 | 0 |
| 0 | -1 | -2 | 0 |

Weight Quantization

```
print("Size of gzipped baseline Keras model: %.2f bytes" % (get_gzipped_model_size(keras_file)))  
print("Size of gzipped pruned Keras model: %.2f bytes" % (get_gzipped_model_size(pruned_keras_file)))  
print("Size of gzipped pruned TFlite model: %.2f bytes" % (get_gzipped_model_size(pruned_tflite_file)))
```

```
Size of gzipped baseline Keras model: 61148.00 bytes  
Size of gzipped pruned Keras model: 21098.00 bytes  
Size of gzipped pruned TFlite model: 19437.00 bytes
```

```
INFO:tensorflow:Assets written to: /tmp/tmp8f26o3p6/assets  
INFO:tensorflow:Assets written to: /tmp/tmp8f26o3p6/assets  
WARNING:absl:Buffer deduplication procedure will be skipped when flatbuffer library is not properly loaded  
Saved quantized and pruned TFlite model to: /tmp/tmp8f26o3p6/assets  
Size of gzipped baseline Keras model: 61148.00 bytes  
Size of gzipped pruned and quantized TFlite model: 6632.00 bytes
```

+ Optimisation du modèle

Pruning

```
print("Size of gzipped baseline Keras model: %.2f bytes" % (get_gzipped_model_size(keras_file)))  
print("Size of gzipped pruned Keras model: %.2f bytes" % (get_gzipped_model_size(pruned_keras_file)))  
print("Size of gzipped pruned TFlite model: %.2f bytes" % (get_gzipped_model_size(pruned_tflite_file)))  
  
Size of gzipped baseline Keras model: 61148.00 bytes  
Size of gzipped pruned Keras model: 21098.00 bytes  
Size of gzipped pruned TFlite model: 19437.00 bytes
```

De 61Ko à 19Ko, soit 60% de compression

Weight Quantization

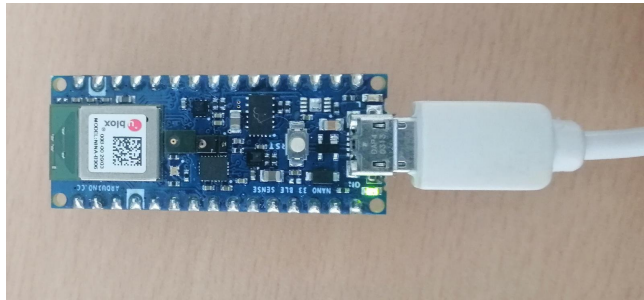
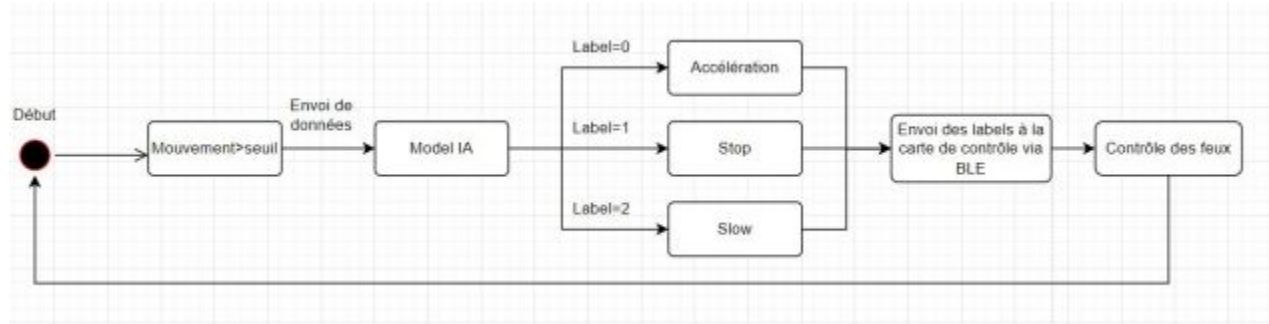
```
INFO:tensorflow:Assets written to: /tmp/tmp8f26o3p6/assets  
INFO:tensorflow:Assets written to: /tmp/tmp8f26o3p6/assets  
WARNING:absl:Buffer deduplication procedure will be skipped when flatbuffer library is not properly loaded  
Saved quantized and pruned TFlite model to: /tmp/tmp8f26o3p6/assets  
Size of gzipped baseline Keras model: 61148.00 bytes  
Size of gzipped pruned and quantized TFlite model: 6632.00 bytes
```

De 19Ko à 6Ko, soit 70% de compression

```
!echo "const unsigned char model[] = {" > /content/model.h  
!cat gesture_model.tflite | xxd -i >> /content/model.h  
!echo "};" >> /content/model.h
```

Le modèle est exporté en fichier .h pour être utilisable sur la carte cible

+ Déploiement du modèle



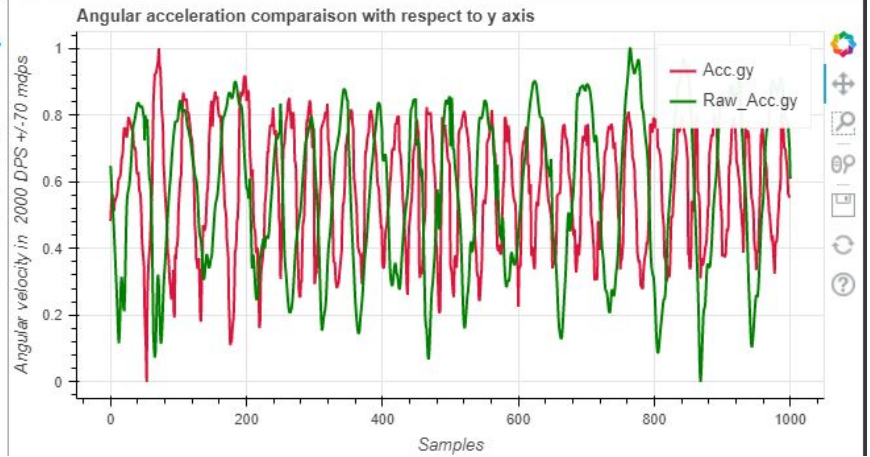
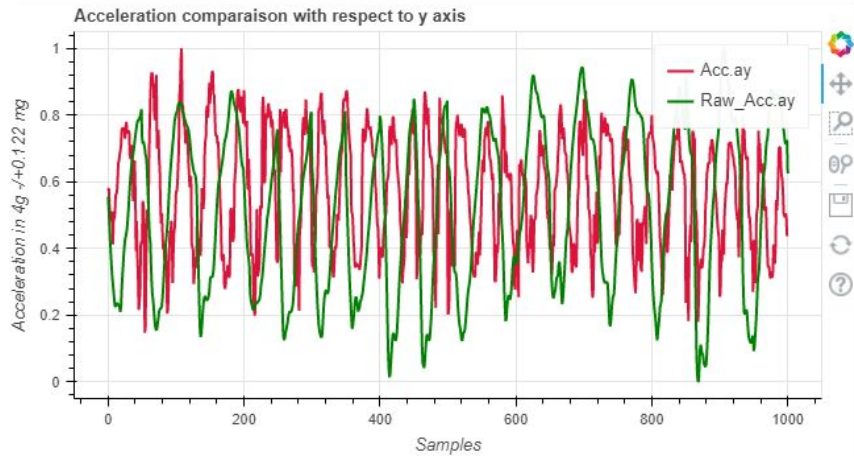
+ Résultats

-> Prédominance de la prédiction du signal d'accélération par rapport aux autres, signifiant une mauvaise précision lorsque le modèle est déployé.

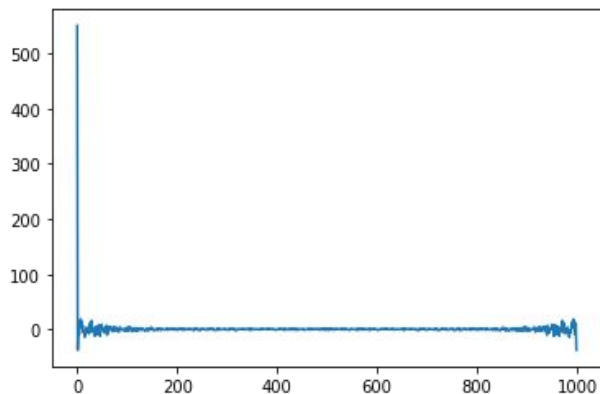
Causes possibles :

- Différence entre les signaux utilisés pendant l'entraînement et le test (Hypothèse testée, ce n'est pas la raison).
- Dégradation du modèle après la quantization ?

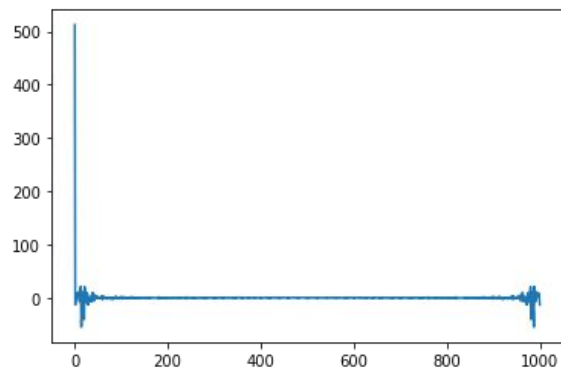
+ Résultats



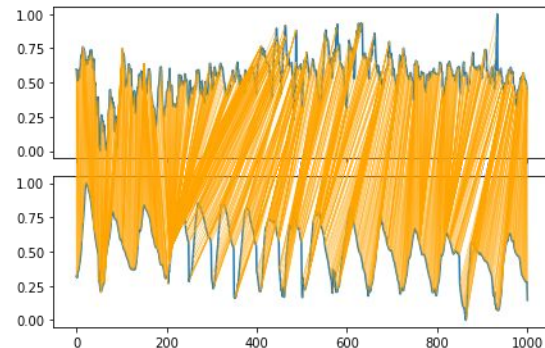
+ Résultats



Spectre du signal d'entraînement

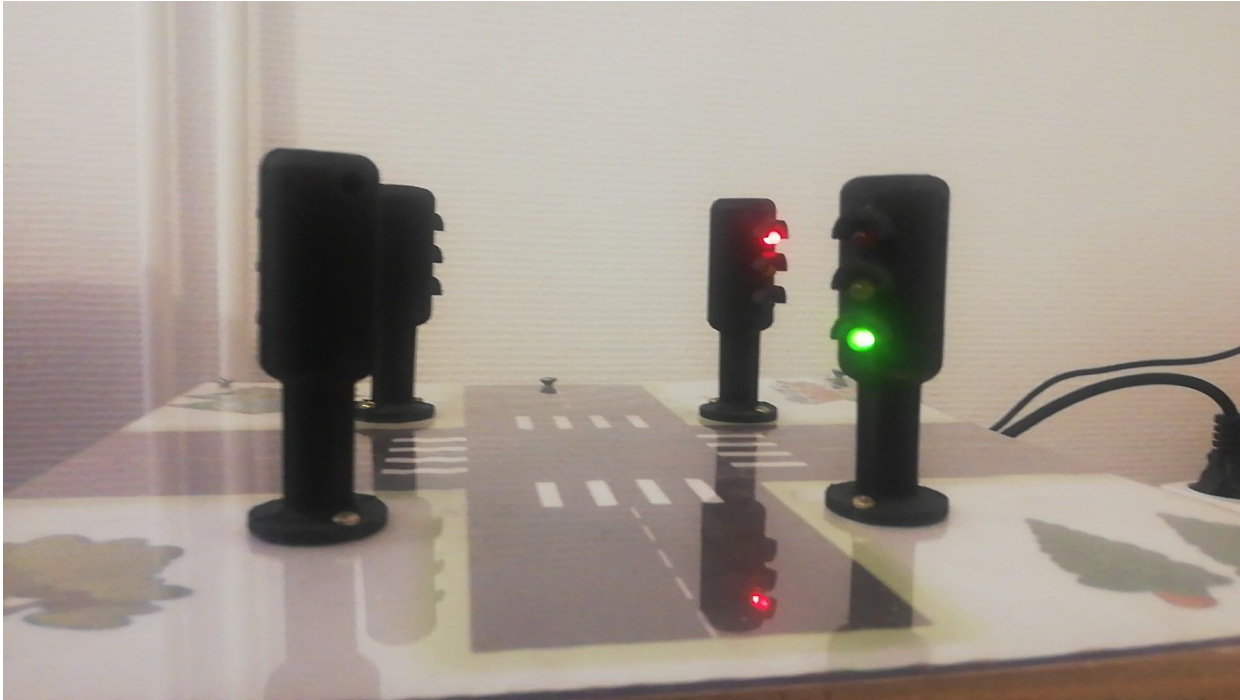


Spectre du signal de test de la même personne



Mesure de similarité entre les deux signaux

+ Démonstration



+ **Conclusion**

The background of the slide features a close-up, slightly blurred image of a guitar's fretboard and strings on the left, and a black microphone on the right. A solid black rectangular box is centered over the image, containing white text.

**Merci pour votre
attention**

Avez vous des questions ?



ENSIM
École d'ingénieurs
Le Mans Université

Présentation du projet 5A :

TinyML pour l'IA embarquée

Réalisé par :

NAIT HAMMOU Mehdi
BEN BRAHIM Abdelkarim

Encadré par :

Pr. HASSAN Kais



Année Universitaire : 2021/2022