

# From Waveforms to Bits

Diogo Fonseca

Professor(s): José Valente de Oliveira & Hélder Daniel

UALG  
MESTRADO EM ENGENHARIA INFORMÁTICA

10th of December, 2025

# Table of contents

## 1 Problem Statement

## 2 Preprocessing and Feature Extraction

## 3 Classifier Model

- Random Forest
- Gradient Boost
- One Vs Rest
- Neural Network
- Stacking Ensembler

# Table of contents

4 Training

5 Results

## Brief summary

## Brief summary

- Identify faults in a physical system.

## Brief summary

- Identify faults in a physical system.
- Classification problem.

## Brief summary

- Identify faults in a physical system.
- Classification problem.
- 7 classes.

## Brief summary

- Identify faults in a physical system.
- Classification problem.
- 7 classes.
- 140 000 samples total.

## Brief summary

- Identify faults in a physical system.
- Classification problem.
- 7 classes.
- 140 000 samples total.
- 20 000 samples for each class.

## Brief summary

- Identify faults in a physical system.
- Classification problem.
- 7 classes.
- 140 000 samples total.
- 20 000 samples for each class.
- data is mostly abstract to humans (sound waves).

# Processing and Feature Extraction

- Shuffling.

# Processing and Feature Extraction

- Shuffling.
- Labels extracted from one\_hot encoding.

# Processing and Feature Extraction

- Shuffling.
- Labels extracted from one\_hot encoding.
- Data split into training and testing.

# Processing and Feature Extraction

- Shuffling.
- Labels extracted from one\_hot encoding.
- Data split into training and testing.
- Data was centered and scaled (mean of 0 and std dev of 1)

# Processing and Feature Extraction

- Shuffling.
- Labels extracted from one\_hot encoding.
- Data split into training and testing.
- Data was centered and scaled (mean of 0 and std dev of 1)
- PCA with variance of 95% (no analysis could be done).

# Classifier Model

## Random Forest

## Random Forest

- 200 decision trees.

## Random Forest

- 200 decision trees.
- no depth limitation.

## Random Forest

- 200 decision trees.
- no depth limitation.
- a node must have at least 5 samples to be considered for splitting.

## Random Forest

- 200 decision trees.
- no depth limitation.
- a node must have at least 5 samples to be considered for splitting.
- each leaf node must have at least 2 samples.

## Gradient Boost

## Gradient Boost

- 200 iterations max.

## Gradient Boost

- 200 iterations max.
- learning rate of 0.1.

## Gradient Boost

- 200 iterations max.
- learning rate of 0.1.
- regularization at 0.1.

## Gradient Boost

- 200 iterations max.
- learning rate of 0.1.
- regularization at 0.1.
- early stopping.

## One Vs Rest

## One Vs Rest

- neural network.

## One Vs Rest

- neural network.
- 256-128 topology.

## One Vs Rest

- neural network.
- 256-128 topology.
- relu activation function.

## One Vs Rest

- neural network.
- 256-128 topology.
- relu activation function.
- daptive learning rate.

## One Vs Rest

- neural network.
- 256-128 topology.
- relu activation function.
- daptive learning rate.
- regularization of 0.0001.

## One Vs Rest

- neural network.
- 256-128 topology.
- relu activation function.
- daptive learning rate.
- regularization of 0.0001.
- 100 iterations max.

## One Vs Rest

- neural network.
- 256-128 topology.
- relu activation function.
- daptive learning rate.
- regularization of 0.0001.
- 100 iterations max.
- early stopping.

## Single Neural Network

## Single Neural Network

- 512-256 topology.

## Single Neural Network

- 512-256 topology.
- relu activation function.

## Single Neural Network

- 512-256 topology.
- relu activation function.
- daptative learning rate.

## Single Neural Network

- 512-256 topology.
- relu activation function.
- daptive learning rate.
- regularization of 0.0001.

## Single Neural Network

- 512-256 topology.
- relu activation function.
- daptative learning rate.
- regularization of 0.0001.
- 100 iterations max.

## Single Neural Network

- 512-256 topology.
- relu activation function.
- daptive learning rate.
- regularization of 0.0001.
- 100 iterations max.
- early stopping.

## Stacking Ensembler

## Stacking Ensembler

- logistic regression.

## Stacking Ensembler

- logistic regression.
- inputs of previous models.

# Training

All training was done multi-threaded on 16 CPU cores.

Classifier model	Time (s)
Random Forest	126
Gradient Boost	26
One Vs Rest	33
Neural Network	38
Stacking Ensembler	1

Table 1: Model training time.

# Results

<b>Model</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>f1-score</b>
Random Forest	0.82	0.82	0.82	0.81
Gradient Boost	0.87	0.87	0.87	0.87
One Vs Rest	0.91	0.91	0.91	0.91
Neural Network	0.90	0.90	0.90	0.90
Stacking Ensembler	0.94	0.94	0.94	0.94

Table 2: Model testing metrics.

# Results

## Confusion Matrix

