Severe Epilepsy Classification

● *Pitch deck presentation - GROUP 6*

Viola Cavedoni (mat.866451)
Andrea Palmieri (mat. 921785)

v.cavedoni@campus.unimib.it
a.palmieri13@campus.unimib.it

Course of Signal and Imaging Acquisition and Modeling in Healthcare

April 4th 2024

Table of contents:

PITCH DECK PRESENTATION

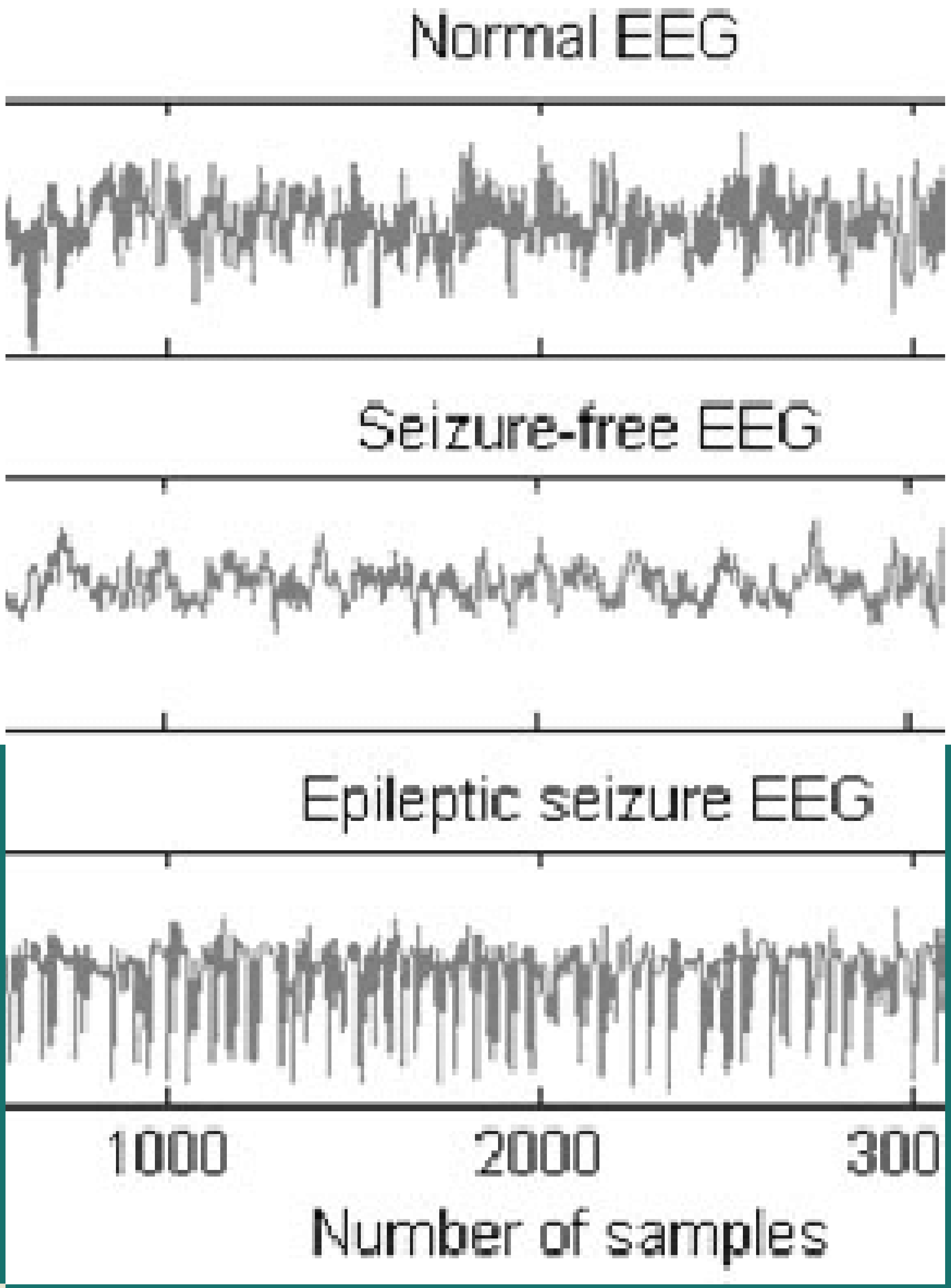
Epilepsy

A little bit of context

Epilepsy is a chronic noncommunicable disease of the brain, characterized by recurrent episodes of involuntary movement (**seizures**). Those episodes are a result of excessive electrical discharges in a group of brain cells.

Epilepsy is **affecting** about **50 million** people worldwide, and an estimated **5 million** people are **diagnosed** with epilepsy each year. An abnormal electroencephalography (EEG) pattern is a consistent predictor of the disease, which cause is unknown in about 50% of cases. An estimated **25%** of epilepsy cases are potentially **preventable**, and up to **70%** of people living with epilepsy could become **seizure free** with use of appropriate medicines.

Source: https://www.who.int/news-room/fact-sheets/detail/epilepsy



Aim of our product

EEG SIGNALS

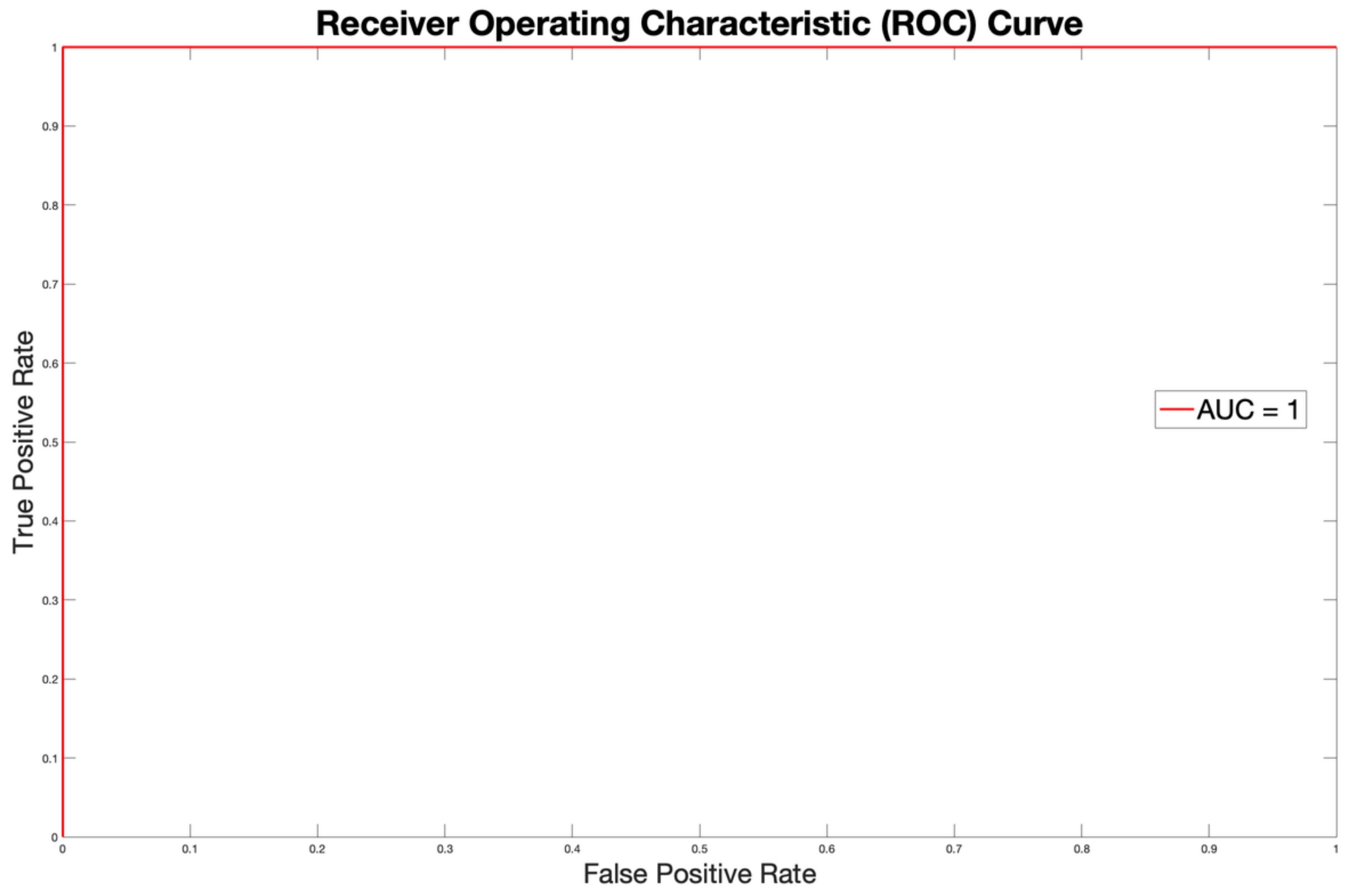
EEGs are signals of the electrical activity of neurons, measured by electrodes placed on the scalp.
Neurons generate potentials which we can model as dipoles, and thanks to the pyramidal cells (aligned perpendicularly to the scalp surface) present in the neocortex, the sum of those dipoles can be measured. We find different types of waves in different neurological conditions.

MISSION

Epilepsy severity levels range from 1 to 5, with 5 being the most severe.

Our aim is to **identify severe (class 4) epilepsy patients** with respect to mild severe (class 3) epilepsy patients given a EEG signal.

Proposed Solution



SVM with linear kernel, BoxConstraint=1, KernelScale=1

TPR ± std	FPR ± std	Accuracy ± std
99.9 % ± 0.1	2.5 % ± 3.1	98.8 % ± 1.5

Frequency Domain Mastery

Our model unlocks deeper insights by transforming EEG signals from discrete time to frequency domain, enhancing diagnostic precision.

Our best model

Our groundbreaking solution harnesses the power of a linear Support Vector Machine (SVM) culminating in unparalleled accuracy and reliability.

Our strengths

High
True Positive Rate
Maximising Sensitivity

Low
False Positive Rate
Minimizing False Alarms

Dataset

200

PATIENTS

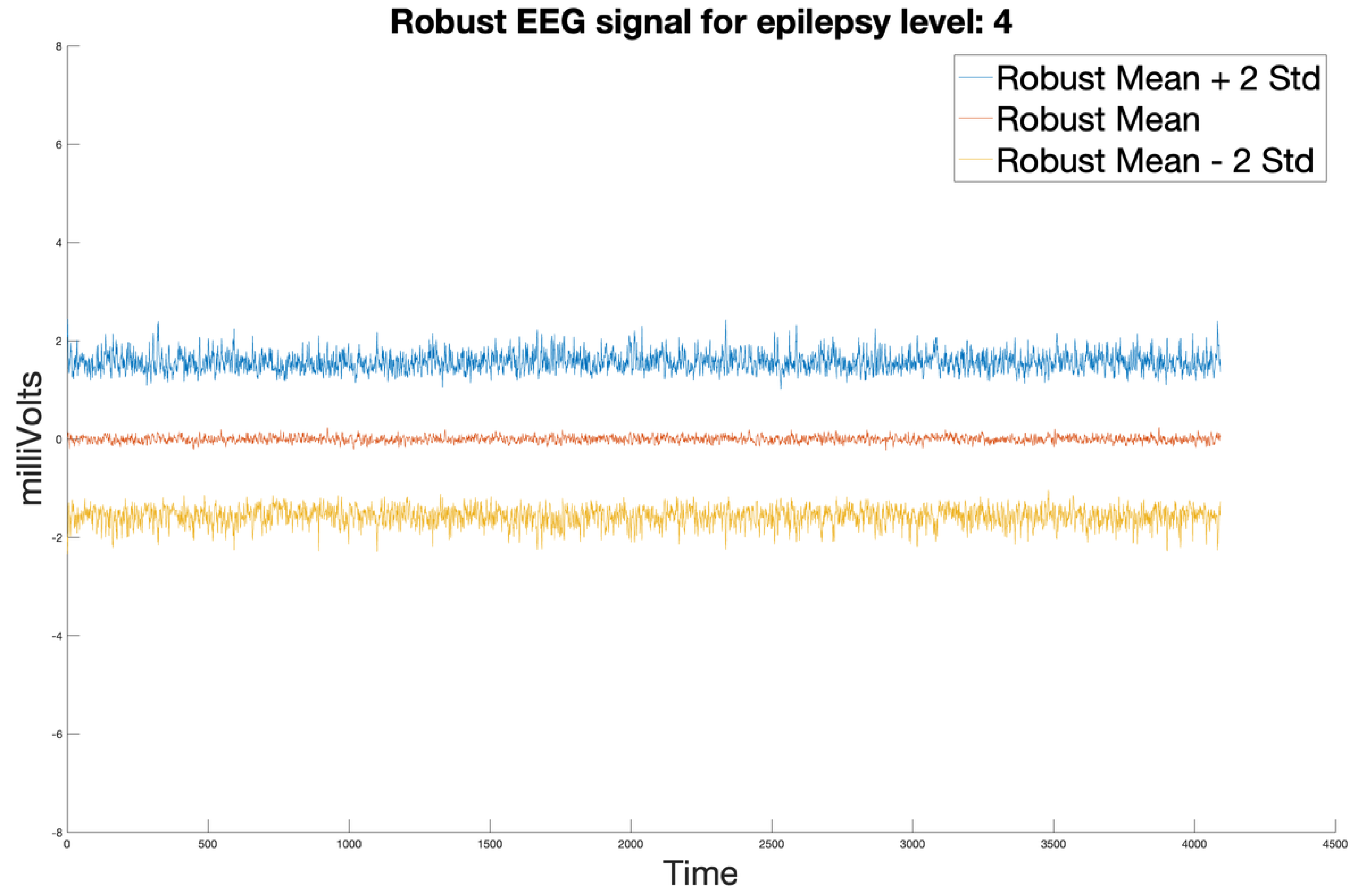
A set of EEG signals of 500 patients with 5 levels of epilepsy was given. The original dataset was reduced to **only patterns of class 3 and 4**, then **split into train data (80%) and test data (20%).** In fact, 100 patterns per level were present in the original dataset, each given by 4094 features.

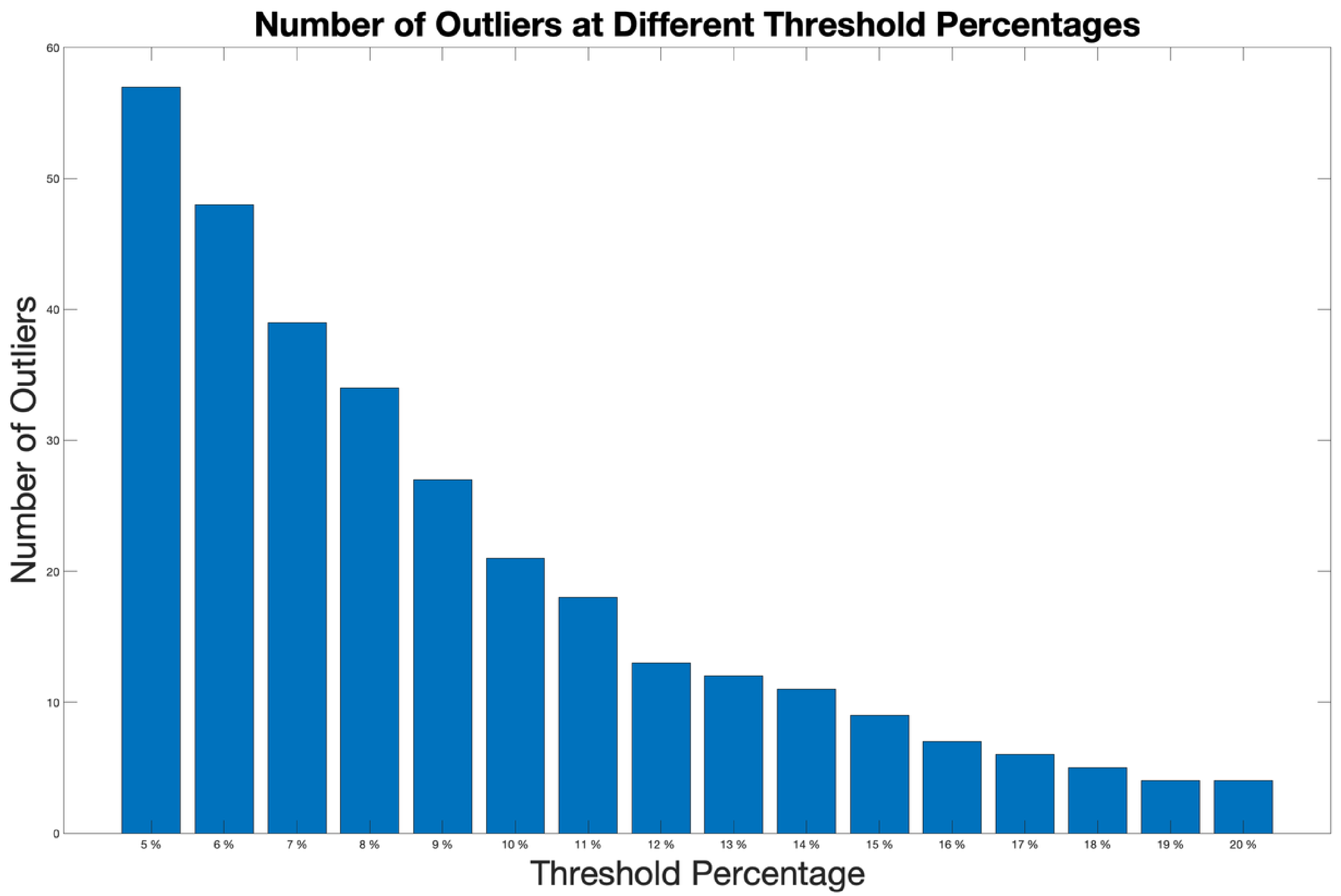
1st	**First approach to training was performed using the original data which is a discrete time signal.**
2nd	**Then second approach to training was performed converting the data to the frequency domain using Fast Fourier Transform.**

Statistical analysis of the dataset

Data were first **robust scaled** using the median and the interquartile range, computed on both class 3 and 4 (as the mean and the standard deviation). Then, outliers detection on the training set is performed using a threshold of 20% of a pattern’s features out of the range [2*std - mean, 2*std + mean].

$$X' = \frac{X - X_{median}}{IQR}$$





Methods: Support Vector Machine

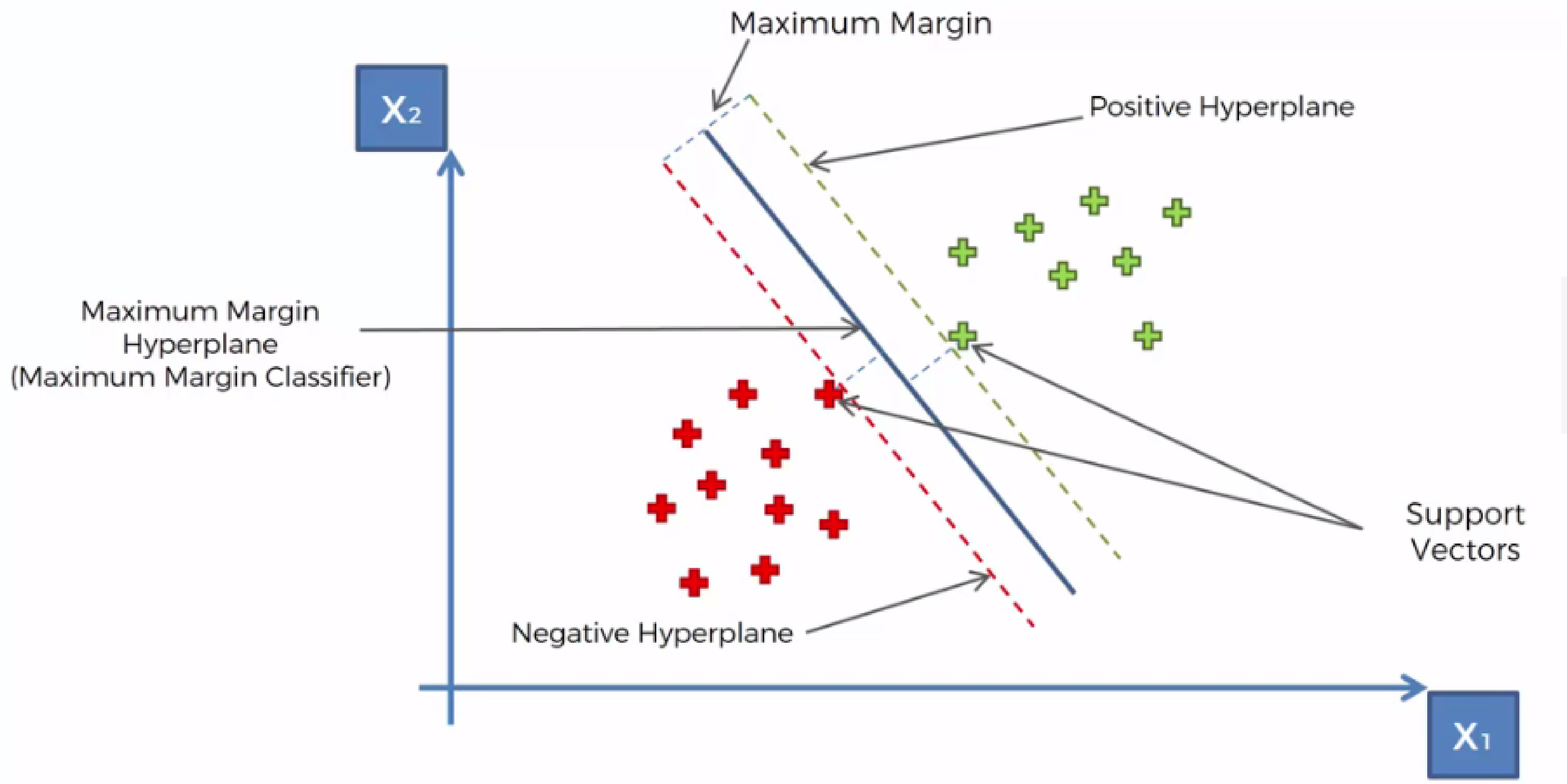
Support Vector Machine is a supervised learning algorithm for classification/regression tasks which works by mapping data to a high-dimensional feature space by a **kernel function**, in order to separate the classes with an **hyperplane**. The model finds marginal lines that define the minimum distance between different class patterns and the hyperplane: samples lying on the margins are called **support vectors**. The wider this margin, the better the model will be at predicting the class of new patterns.

Thus the goal is to find the hyperplane that maximizes the margin.

Some parameters are introduced.

Box constraint controls the penalty imposed on margin-violating samples. Decreasing the box constraint means tolerating more misclassifications. Also **scaling** the kernel can improve the performance because it enlarges the space between the classes.

Finally, the choice of the kernel (either linear, gaussian or polynomial) is important.

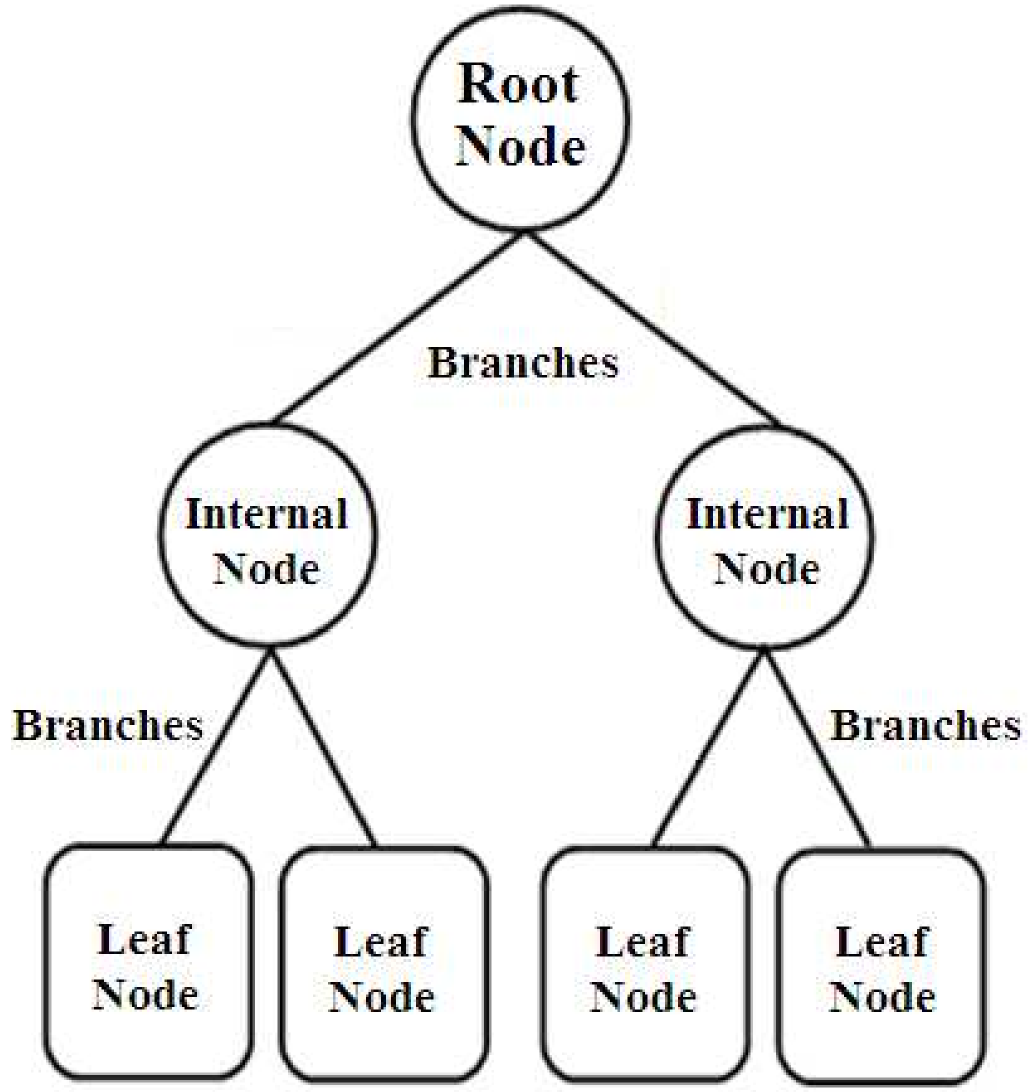


Methods: Decision Trees

A decision tree is a supervised learning algorithm for classification/regression tasks characterized by a **hierarchical** tree **structure**. It starts with a **root node** without any incoming branches. The outgoing branches feed into internal nodes (**decision nodes**). Basing on the available features those nodes perform evaluations to form subsets represented by leaf nodes (**terminal nodes**). The leaf nodes represent all the possible outcomes.

Decision trees employ **"divide and conquer"** strategy to identify the optimal splits. The **splitting process** is repeated until the majority of patterns have been classified.

One parameter is the maximum number of decision splits. Also the **split criterion** is important: for example the **Gini Index** measures the probability for a random chosen sample to be misclassified, and in this case the splits are chosen according to the feature that minimizes the index.



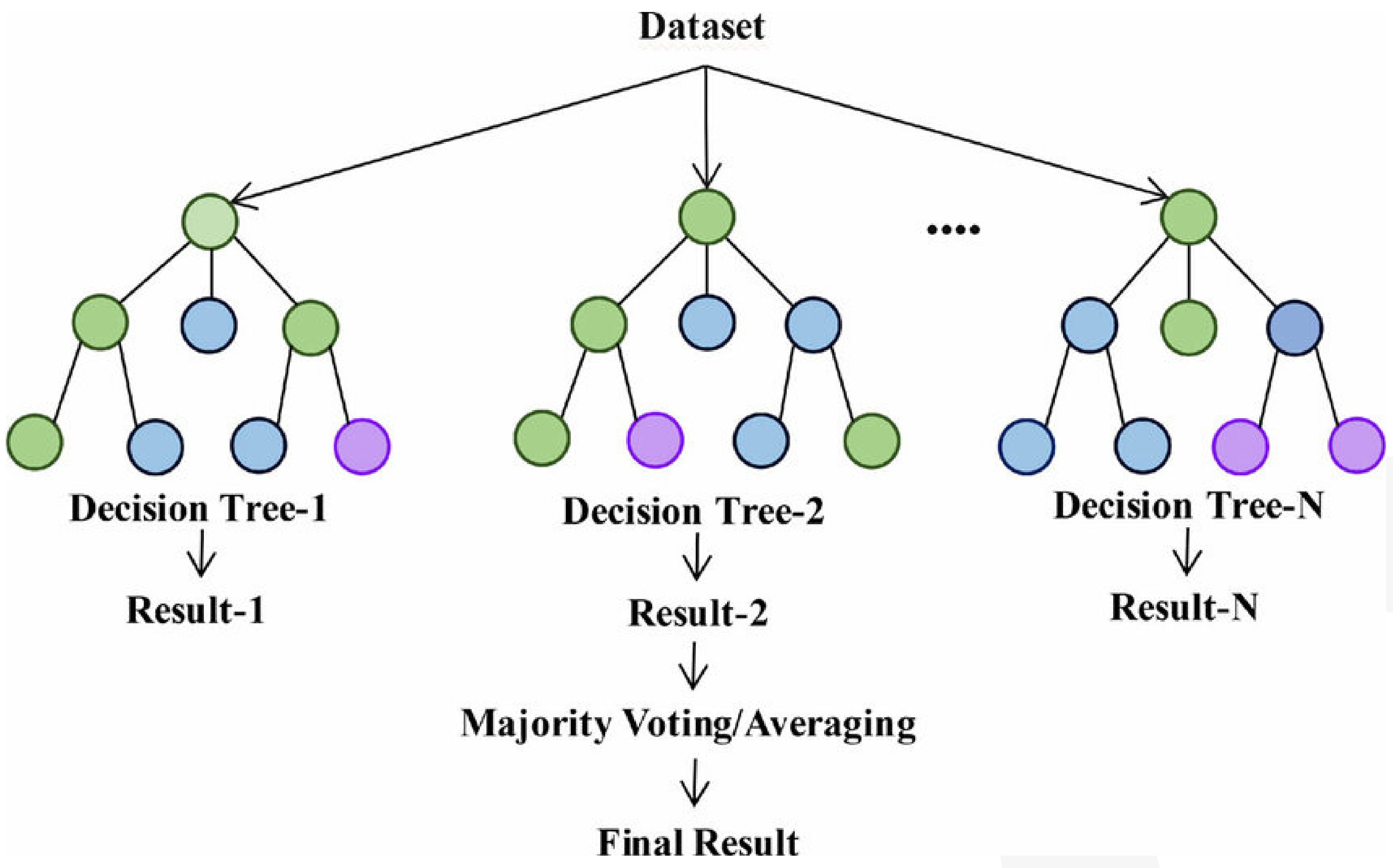
Methods: Random Forest

Random forest is an **ensemble method** for classification/regression tasks. An ensemble method trains a set of learning algorithms (in this case decision trees) and combines their outputs to solve a single problem. In particular random forest it's an extension of the bagging algorithm.

Bagging employs extraction with replacement from the original training set to train the decision trees (this is called bootstrap aggregation) and voting to select the final output.

In addition to this, random forest employs **random feature selection** for every decision tree involved: in this way, the trees will be low correlated.

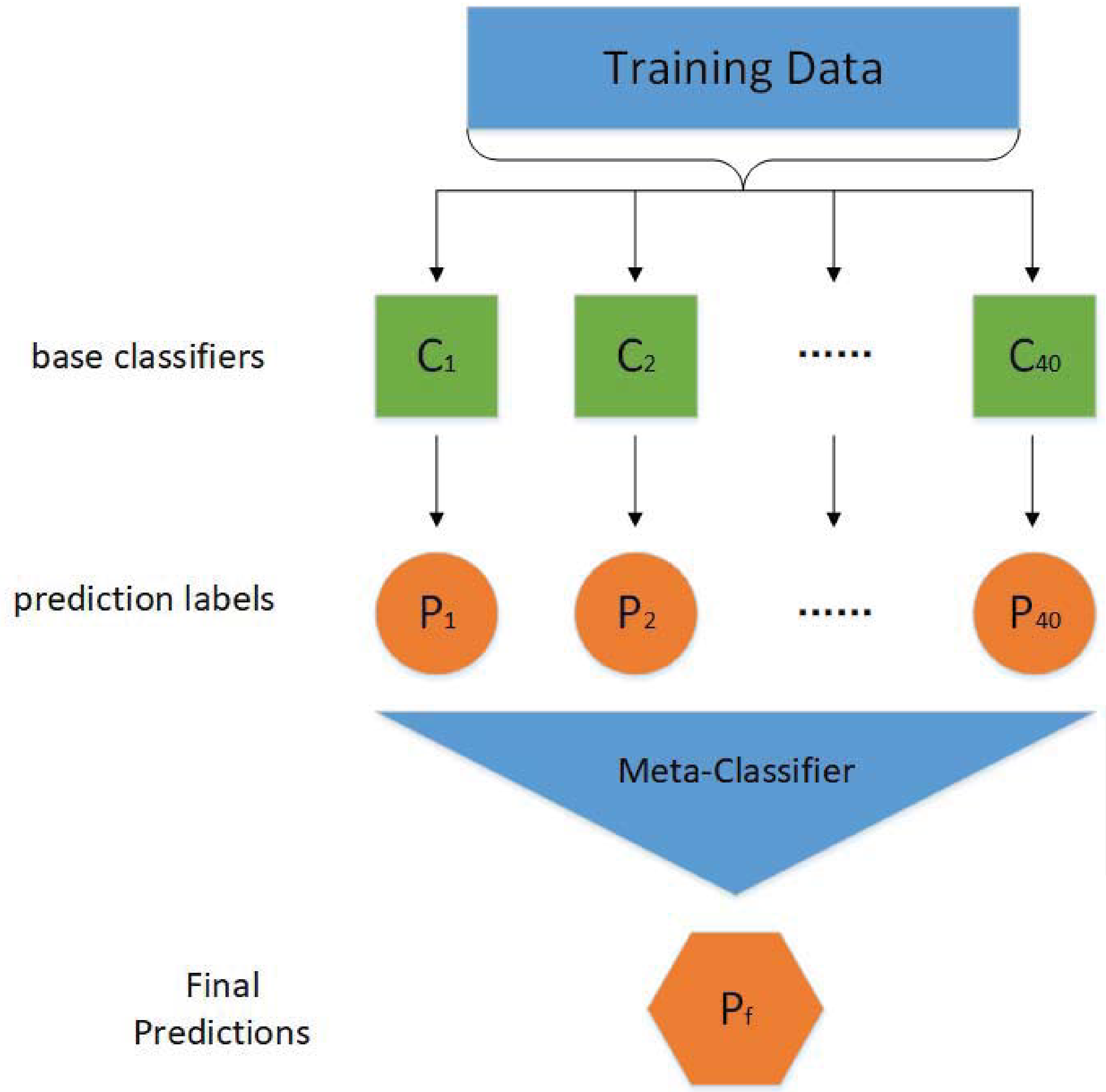
An important parameter is the number of decision trees to be trained.



Methods: Stacked Classifier

Also stacked classifiers are ensemble methods. In particular we find a **first level** of classifiers trained with the original training data.
Their predictions or the corresponding scores (confidence on predictions) are then used to train a second level consisting of a **meta classifier**, which will give the final output.

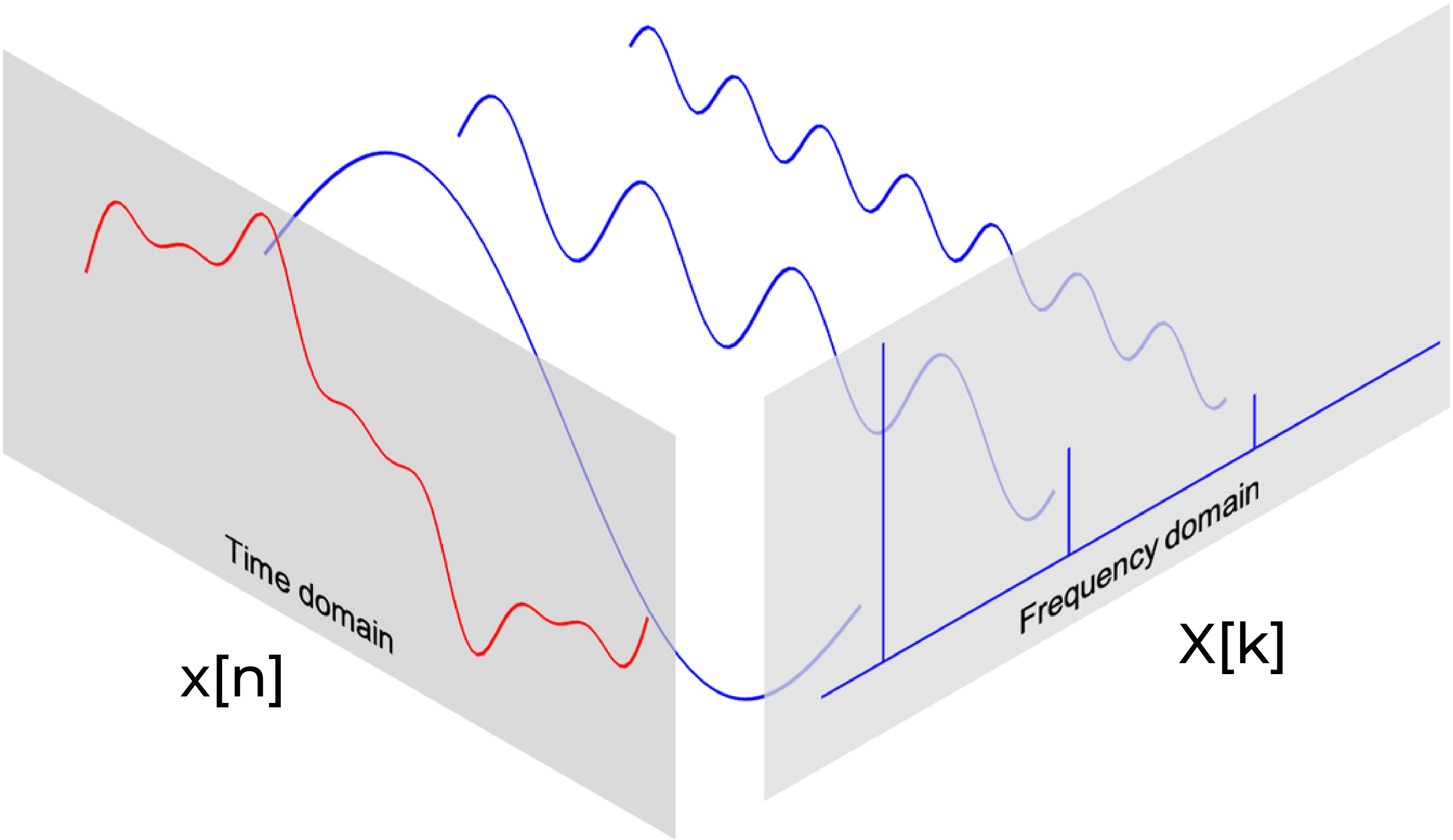
For this project our first level is composed by a SVM with linear kernel, a decision tree and a random forest classifiers, and the meta classifier is again a random forest classifier.



Methods: Discrete Fourier Transform

The **Discrete Fourier Transform (DFT)** allows the transformation of **discrete-time signals x[n]** into **frequency-domain representations X[k].**

This enables us to ***identify patterns*** within EEG data by isolating their **constituent frequency** components.



$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-j\frac{2\pi}{N}kn}$$

DFT formula

Nowadays it’s computed using the
Fast Fourier Transform (FFT) algorithm

Metrics

Confusion Matrix

		Predicted	
		Positive	Negative
Actual	Positive	**True Positive**	**False Negative**
	Negative	**False Positive**	**True Negative**

True Positive Rate

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

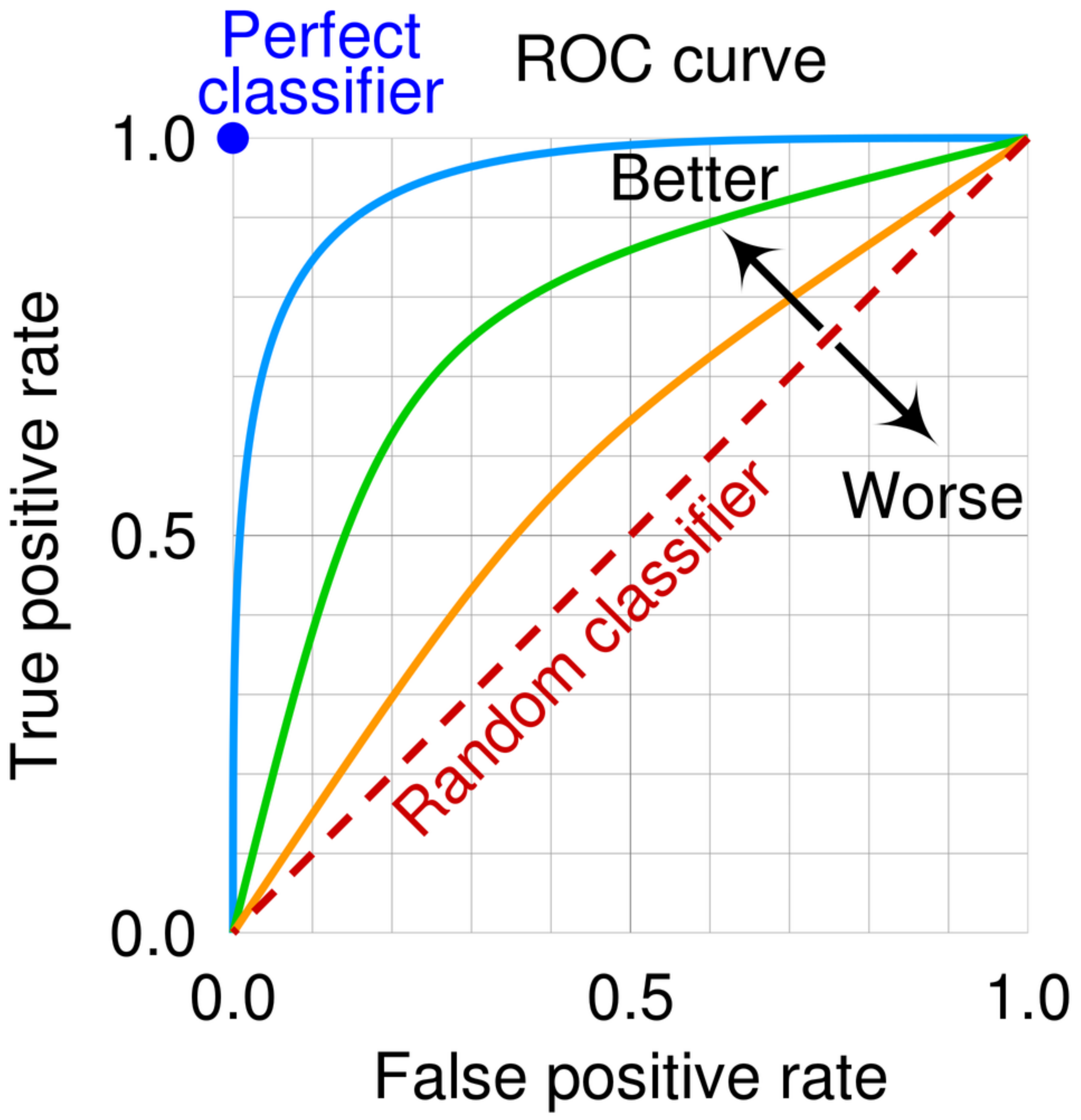
Higher is better

False Positive Rate

$$FPR = \frac{FP}{FP + TN}$$

Lower is better

Receiver Operating Characteristic (ROC)



Results: training on discrete-time signals

Hyperparameters optimization for the **SVM** was performed with a **Grid Search** across a set of possible values and the **best combination** was used for each kernel type.

Training metrics are obtained by performing a **k=5 cross-validation.**

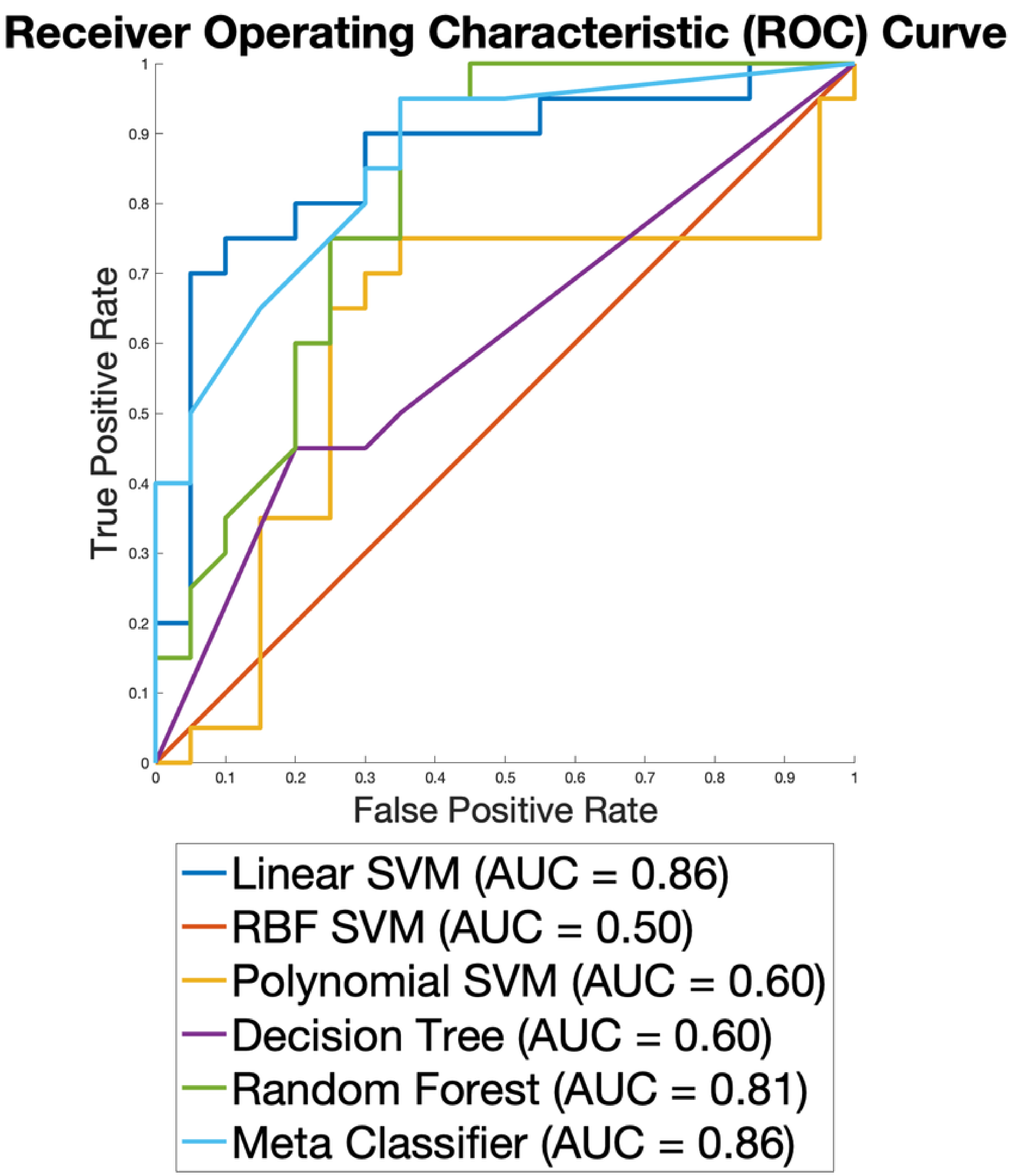
	TPR ± std	FPR ± std	Accuracy ± std
SVM gaussian kernel, BoxConstraint=1, KernelScale=1	100 % ± 0	100 % ± 0	50.6 % ± 1.3
SVM polynomial kernel, BoxConstraint=1, KernelScale=5	92.5 % ± 9.2	97.3 % ± 5.3	48.1 % ± 5.4
SVM linear kernel, BoxConstraint=1, KernelScale=5	54.8 % ± 23.9	15.7 % ± 5.6	69.3 % ± 11.8
Decision Tree, SplitCriterion=gdi, MaxNumSplits=100	54.5 % ± 13.1	39.3 % ± 12.7	57.6 % ± 12.7
Random Forest, 100 trees	69.9 % ± 14.3	23.7 % ± 16.1	73.0 % ± 9.2

Results: testing on discrete-time signals

Testing revealed that **no model was reliable enough** to be used for classification.
The linear SVM has a low FPR (5 %) however a low TPR, with a AUC of 0.86.
The **best model is the meta-classifier** trained on the scores of the level-1 classifiers.

	TPR	FPR	Accuracy	AUC
SVM gaussian kernel, BoxConstraint=1, KernelScale=1	100 %	100 %	50 %	0.50
SVM polynomial kernel, BoxConstraint=1, KernelScale=5	95 %	95 %	50 %	0.61
SVM linear kernel, BoxConstraint=1, KernelScale=5	55 %	5 %	75 %	0.86
Decision Tree, SplitCriterion=gdi, MaxNumSplits=100	50 %	65 %	42.5 %	0.60
Random Forest, 100 trees	80 %	35 %	72.5 %	0.81
Meta-classifier, Random Forest, 100 trees	85 %	30 %	77.5 %	0.86

Level-1 classifiers

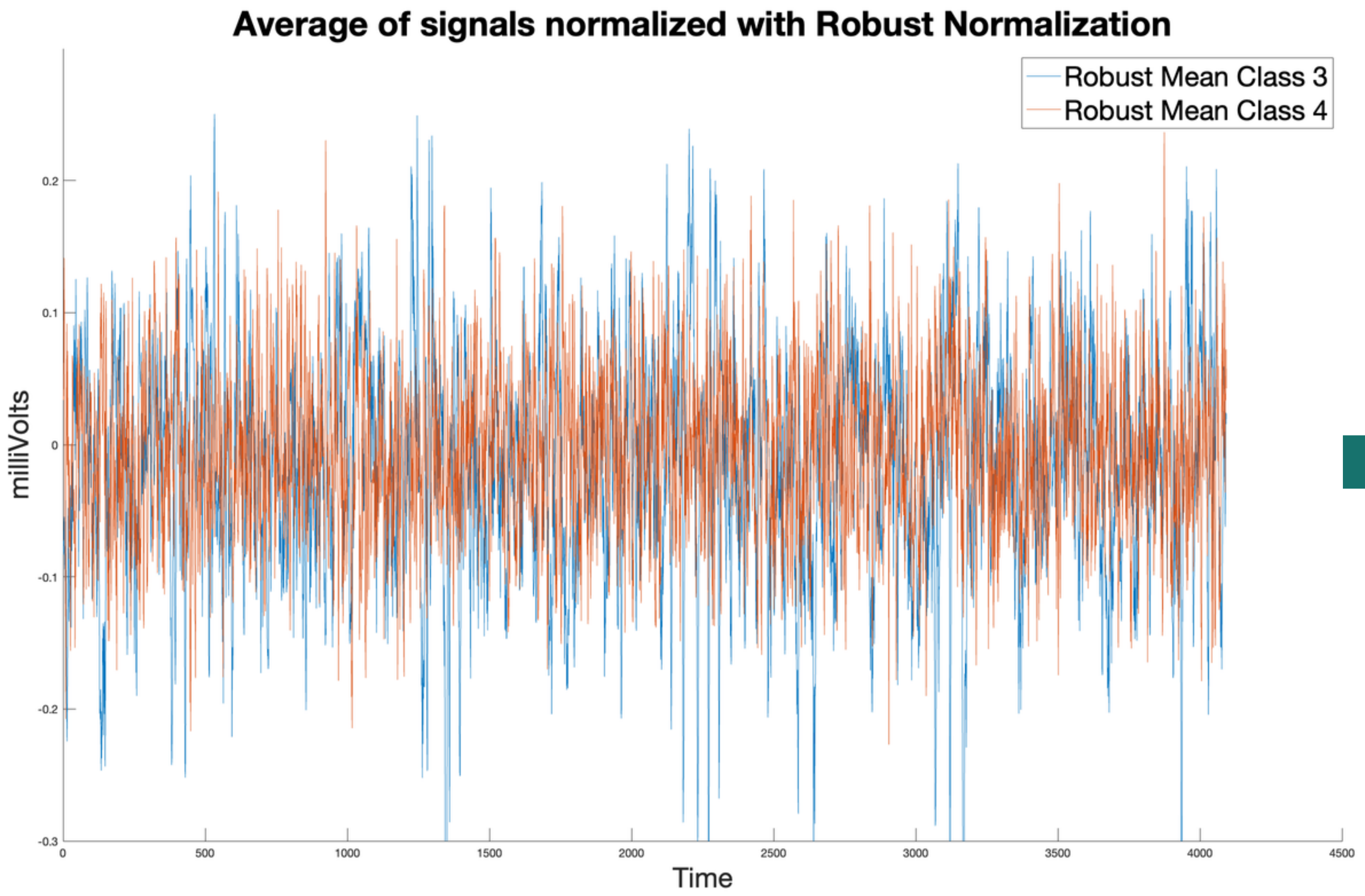


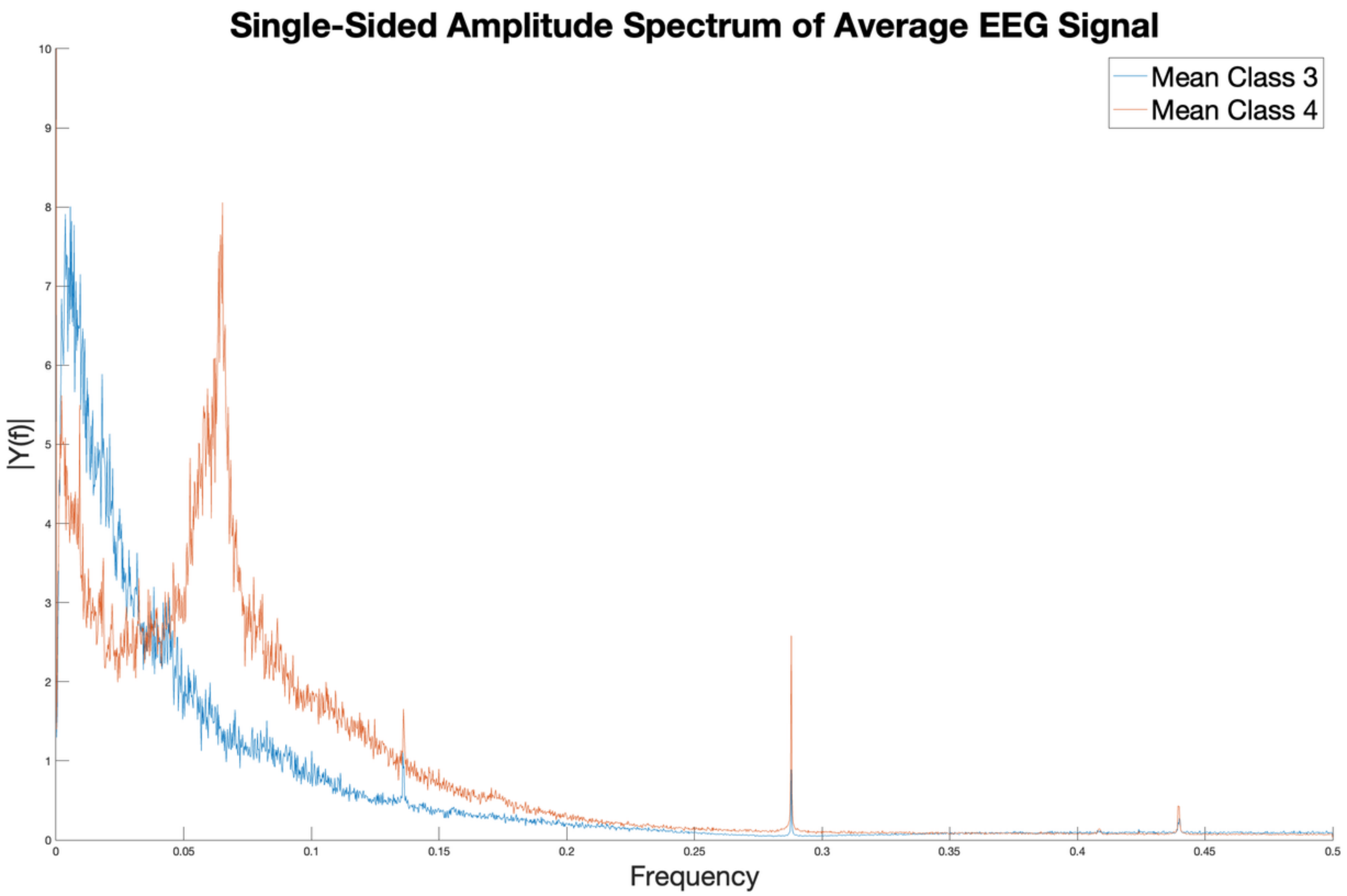
Frequency spectrum

Transforming the EEG signals from discrete-time domain to the frequency-domain **reveals interesting patterns** for each class.

The average signals for each class are shown below on the right:

- **2 distinct peaks** are found





Results: training on frequency-domain

Two models were tested for the frequency-domain:

- **SVM with linear kernel**
- **Decision Tree.**

The following metrics were found performing **k=5 Cross-Validation.**

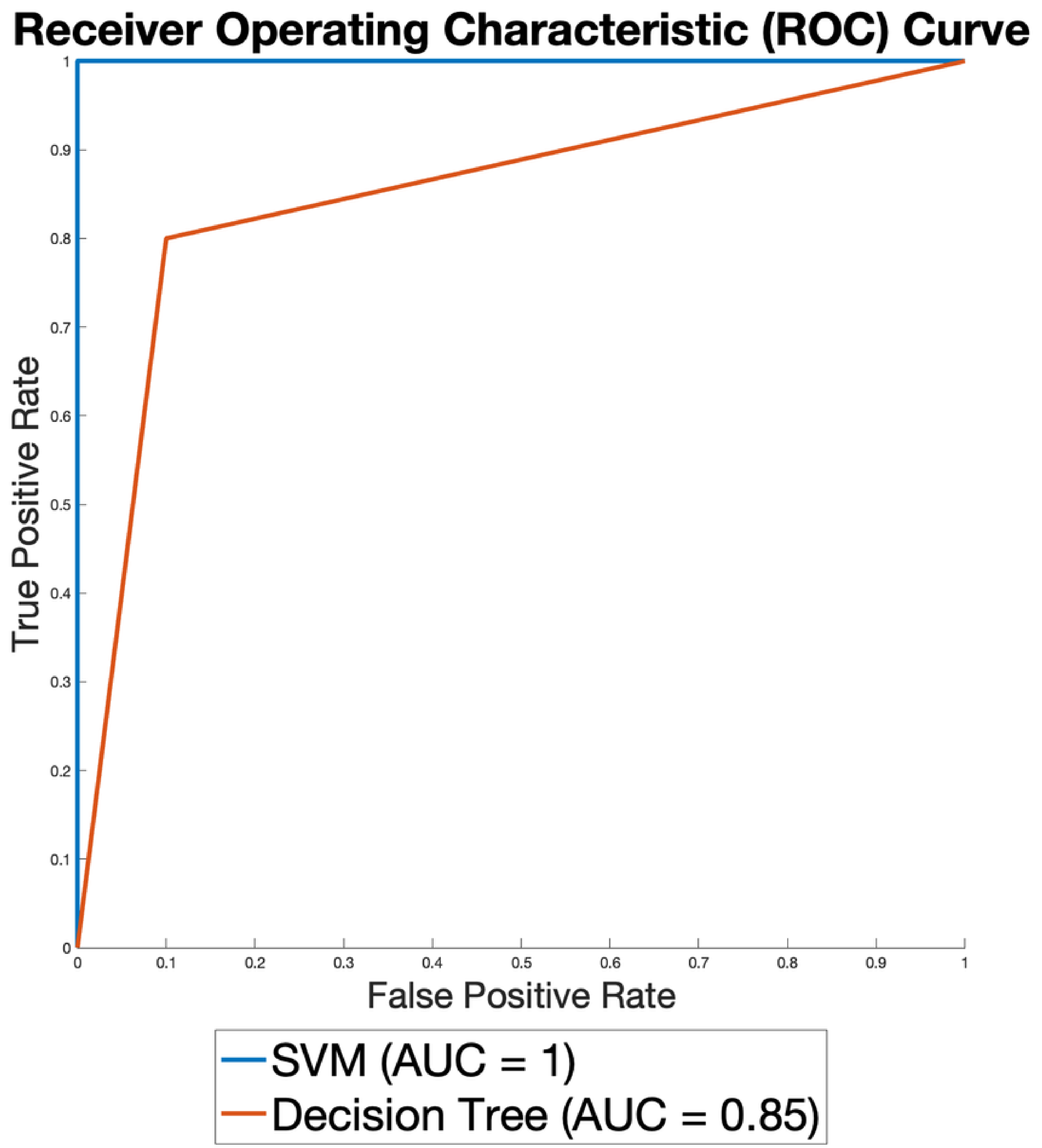
	TPR ± std	FPR ± std	Accuracy ± std
SVM linear kernel, BoxConstraint=1, KernelScale=1	**99.9 % ± 0.1**	**2.5 % ± 3.1**	**98.8 % ± 1.5**
Decision Tree, SplitCriterion=gdi, MaxNumSplits=100	91.3 % ± 7.5	11.3 % ± 10.0	90.0 % ± 3.6

Results: testing on frequency-domain

Comparative results of testing two classifiers on EEG data transformed into the frequency-domain.

SVM excels at classifying correctly the two classes.

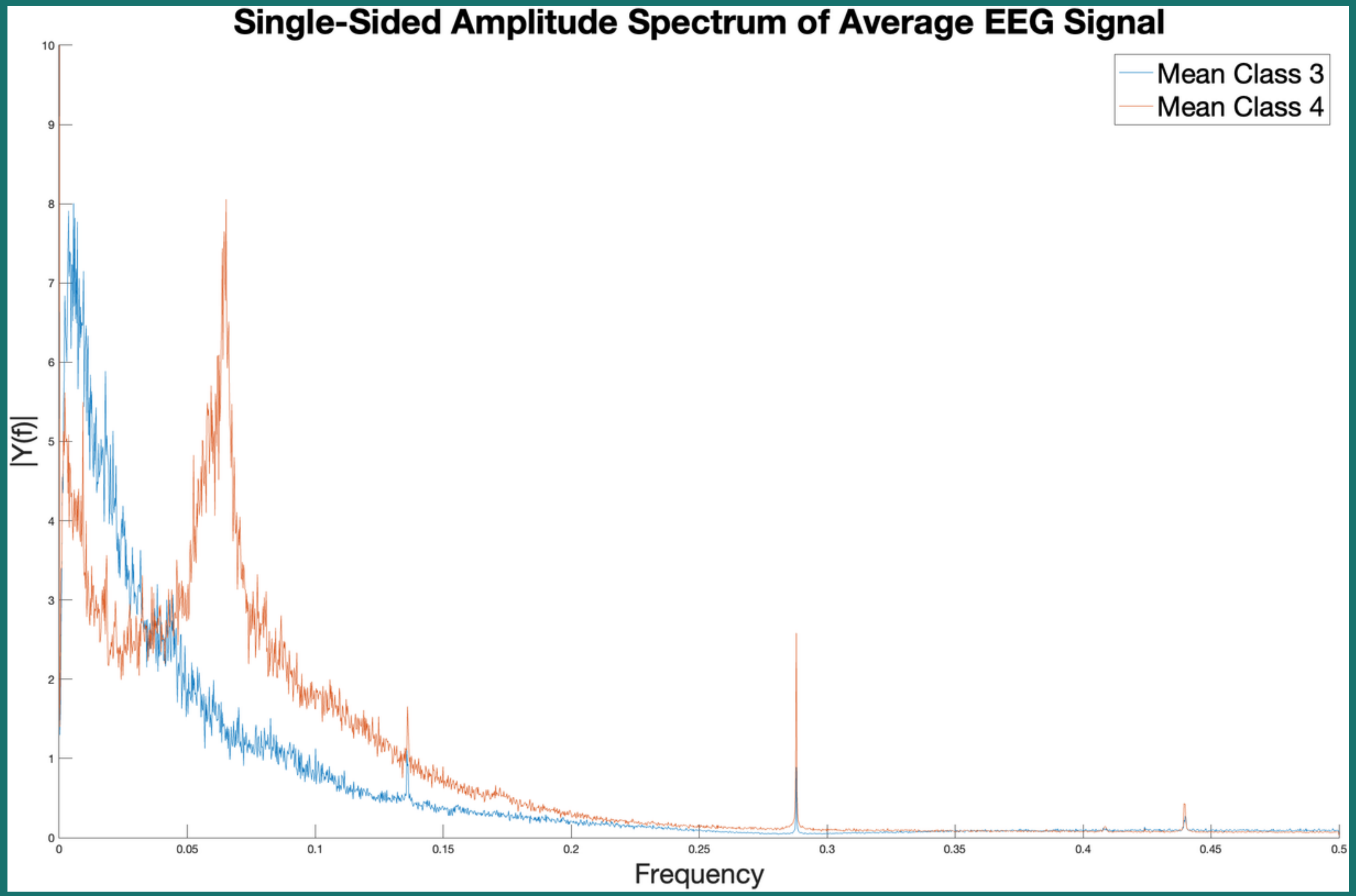
	TPR	FPR	Accuracy	AUC
SVM linear kernel, BoxConstraint=1, KernelScale=1	**100 %**	**0 %**	**100 %**	**1**
Decision Tree, SplitCriterion=gdi, MaxNumSplits=100	80 %	10 %	85 %	0.85



Robustness...

... our motto!

Due to the marked difference between the two classes, our product achieves stellar levels of Accuracy, TPR and FPR.



Emphasizing the Strength of Our Model

The dataset undergoes **20 iterations of data splitting** between training and test sets using different seeds.

Testing is conducted for each iteration, and the table displays the average and standard deviation of metrics across all iterations of data splitting.

TPR ± std	FPR ± std	Accuracy ± std
99.9 % ± 0.1	**0.5 % ± 1.5**	**99.75 % ± 0.8**

SVM with linear kernel, BoxConstraint=1, KernelScale=1

SITOGRAPHY

https://it.mathworks.com/products/matlab.html

www.ibm.com

www.towardsdatascience.com

https://www.who.int/

EEG IMAGE

Waveform of three EEG signals normal (set A), seizure-free (set D) and... | Download Scientific Diagram (researchgate.net)

SVM IMAGE

Arun Manglick - Artificial Intelligence & Machine/Deep Learning: Support Vector Machine (SVM) (arun-aiml.blogspot.com)

DECISION TREE IMAGE

(PDF) Lightning Forecast Using Data Mining Techniques On Hourly Evolution Of The Convective Available Potential Energy (researchgate.net)

RANDOM FOREST IMAGE

(PDF) A comparative evaluation of machine learning algorithms and an improved optimal model for landslide susceptibility: a case study (researchgate.net)

STACKED CLASSIFIER IMAGE

(PDF) An Automated Approach to Diagnose Turner Syndrome Using Ensemble Learning Methods (researchgate.net)

THANK YOU

FOR YOUR ATTENTION

Viola Cavedoni (mat.866451)
Andrea Palmieri (mat. 921785)

v.cavedoni@campus.unimib.it
a.palmieri13@campus.unimib.it

Course of Signal and Imaging Acquisition and Modeling in Healthcare

April 4th 2024