**Multi-Label Classification**

**Scikit-learn Library used : sklearn.multiclass**

**Strategy used: Multi-Label Classification :** Classification task labelling each sample with labels with . Comparable to running binary classification tasks with **sklearn.multioutput.MultiOutputClassifier.** However, this approach treats each label independently whereas multilabel classifiers *may* treat the multiple classes simultaneously, accounting for correlated behavior among them (**may learn the prior in the way pathologies can appear simultaneously**): we will try both type of Classifiers.

We will use a **One-Vs-The-Rest** approach in order to train our models.

**Classifier used**: We are going to use a Gradient Boosting Classifier (implemented and fitted in the way Armand will tell us) in a OneVsRest approach.

The hyperparameters tuned for my model are:

|  |  |  |  |
| --- | --- | --- | --- |
| **Hyperparameter** | **Explanation** | **Value Range** | **Cardinality** |
| Learning Rate |  | [0.0001;0.2] | 10 |
| Max depth | Depth of One tree | [5:10] | 5 |
| Min\_child\_weight | The minimal weight to split nodes | [10:30] | 5 |
| Subsample | Subsample of training data to create a tree | [0.6; 0.9] | 5 |
| Col\_sample\_by\_tree | Subsample of features used by tree | [0.6; 0.9] | 5 |
| Col\_sample\_bylevel | Subsamples used for each level of a tree | [0.6; 0.9] | 5 |
| Reg\_alpha | L2 Regularization term | [1; 5] | 5 |
| Reg\_lambda | L1 Regularization term | [1; 5] | 5 |
| Num\_parallel\_tree | Number of parallel trees constructed | [200; 800] | 10 |
| Scale\_pos\_weight | Class weighting | #Negative/#Positive |  |

**Data Format:** The format of label expected is an array of size with, for every sample, 1 in the position where the corresponding label is positive, else 0.

**Training/Testing split:** We will use the same Train/Test split as we have done previously (the one from Janmajay split). The Training Set is ready. However, for the testing set I need to reextract features in order to account for multi-labelled examples. Furthermore, I will lower the number of SNR examples in the testing set in order to add them to the training Set. (It should take me 24 hours to extract the new features for the Testing Set).

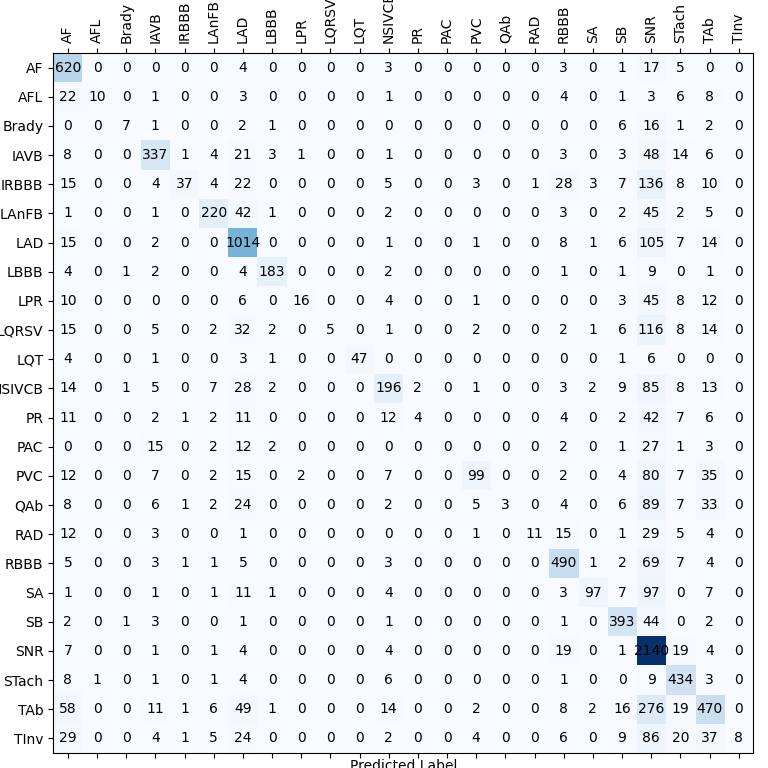
I will perform Cross Fold Validation on the training Base and evaluate my final model on the Testing Base. If the feature Selection process is finished by the time I start these experiments, I will perform these experiments with only the features selected.

**Experiments to be done:** How to insert prior on the way pathologies may appear simultaneously?

**Results:**

**Failure in Hyperparameter tuning**: the tuning of hyperparameter takes too much time and raises some unexpected errors. I performed this experiment with a default RandomForestClassifier and XGBoost. I got better results using Random Forest, which is not using weighting for now.

**Global Competition score on Test Set:** 0.62 (**Minimum preprocessing : Bandpass 0.05-100Hz+ Notch [Bosser dès maintenant: [0.05-100 + Notch à 50 et 60]**)

**Fbeta scores on Test Set:** {'AF': 0.9127584297732705, 'AFL': 0.19147084421235855, 'Brady': 0.18461538461538463, 'IAVB': 0.7241416647443473, 'IRBBB': 0.14489227691725465, 'LAnFB': 0.5986342835521478, ' LAD': 0.7193715897579337, 'LBBB': 0.8034053700065487, 'LPR': 0.17584480600750935, 'LQRSV': 0.03473099267491794, 'LQT': 0.7469879518072289, 'NSIVCB': 0.5864785079732936, 'PR': 0.03551136363636363, 'PAC': 0.0, 'PVC': 0.42356494155477353, 'QAb': 0.01700654845558114, 'RAD': 0.13387544625148756, 'RBBB': 0.8026856258146939, 'SA':

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Pathologies** | **Number of Examples**  **In the Competition DataBase** | **Number of Features extracted specifically for this pathology per lead** | **Fbeta score**  **Multi-Label Classification** | **Fbeta score**  **Single-Label Classification** | **Opportunities for Data Augmentation** |
| AF | 2345 | 16 | 0.91 | 0.84 |  |
| AFL | 308 | 16 (the ones from AF) | 0.20 | 0.25 |  |
| Brady | 259 | 1 | 0.20 | 0.26 |  |
| IAVB | 1318 | 3 | 0.72 | 0.61 |  |
| IRBBB | 1221 | 0 | 0.14 | 0.02 |  |
| LAnFB | 1254 | 0 | 0.60 | 0.42 |  |
| LAD | 2126 | 2 | 0.72 | 0.57 |  |
| LBBB | 982 | 20 | 0.80 | 0.85 |  |
| LPR | 338 | 6 (3 AVB, 3 challenge) | 0.17 | 0.11 |  |
| LQRSV | 526 | 5 | 0.03 | 0.15 |  |
| LQT | 1090 | 5 (challenge) | 0.75 | 0.33 |  |
| NSIVCB | 897 | 0 | 0.60 | 0.04 |  |
| PR | 299 | 0 | 0.03 | 0.51 |  |
| PAC | 1337 | 12 | 0.0 | 0.76 |  |
| PVC | 552 | 26 | 0.42 | 0.18 |  |
| QAb | 824 | 5 | 0.68 | 0.01 |  |
| RAD | 403 | 2 | 0.79 | 0.22 |  |
| RBBB | 2018 | 20 (the ones from LBBB) | 0.80 | 0.70 |  |
| SA | 1087 | 0 | 0.44 | 0.50 |  |
| SB | 1606 | 0 | 0.80 | 0.64 |  |
| SNR | 12019 | 0 | 0.80 | 0.80 |  |
| STach | 1555 | 0 | 0.90 | 0.82 |  |
| TAb | 1865 | 5 | 0.46 | 0.14 |  |
| Tinv | 832 | 5 | 0.80 | 0.03 |  |

**Color code for the Table:**

Pathologies for which the score is satisfying: AF, LBBB, RBBB, SBR, SB, STach

Pathologies that have consequently benefitted from the Multi-Label Classification: IAVB, LAnFB, LAD, LQT, NSIVCB, PVC, SB, TAb, Stach, RBBB

Pathologies that have consequently suffered from the Multi-Label Classification: LQRSV, PR, PAC

Pathologies fo which Data-Augmentation is a valuable option: AFL, Brady, LPR, LQRSV, PR, PVC, RAD, TInv (pathologies for which I will first extract new features before Augmenting Data)

Pathologies for which I will extract new features: AFL (in order to separate with AF), Brady, IRBBB, LAnFB, LAD, NSIVCB, PR, QAb, RAD, TAb, SA, TInv

**Repetitive mistakes of our model from the Confusion Matrix:**

Misclassification of pathologies with SNR, AF and LAD: the **three most common pathologies**. The steps we are going to execute in order to prevent such behavior are: **Extract features** for these pathologies if we do not do it yet, **Class weighting** and **Data Augmentation** and **change multilabel strategy ? (cost function factorization) + tune HyperParameters with multilabels.**

**Specificities on Features extracted**

**1. Features extracted for AFL:** I have tried the f\_wave detection (with the frequency of the maximum of the ecg, this did not successfully visually separated AF from AFL). However, I will still use it as a feature and see how well it combines with other features.

**2. Features extracted for IAVB:** The ones I extracted from the paper: *Automated Detection of First-Degree Atrioventricular Block Using ECGs*

*Luning Mao, Hao Chen, Jiaqi Bai, Jieying Wei, Qiang Li, and Rui Zhang*, giving me some 0.85 Fbeta during the unofficial phase

**3. Features extracted for RBBB/LBBB:** Paper : *Automatic Detection of Strict Left Bundle Branch Block Radovan Smisek , Pavel Jurak , Ivo Viscor, Josef Halamek , Filip Plesinger , Magdalena Matejkova, Pavel Leinveber , and Jana Kolarova,* giving me 0.75 Fbeta for RBBB and 0.85 Fbeta for LBBB

**4. Features extracted for PAC:** Paper : *Automatic detection of premature atrial contractions in the electrocardiogram Vessela T. Krasteva, Irena I. Jekova, Ivaylo I. Christov*, giving me 0.80 Fbeta score during the unofficial phase

**5. Features extracted for PVC:** Papers : *Automatic Identification of Premature Ventricular Contraction Using ECGs Hao Chen, Jiaqi Bai, Luning Mao, Jieying Wei, Jiangling Song, and Rui Zhang and PVC discrimination using the QRS power spectrum and self-organizing maps M.L. Talbi ∗, A. Charef*. These features gave me 0.70 Fbeta score in the unofficial phase

**Specificities on the features to be Extracted for new pathologies:**

**1. Features I will extract for AFL [errors: Misclassification with AF]:** The maxfreq from the 12-leads ECG, computed from the functions @Joachim Behar sent me. This sould help to differentiate between AF and AFL according to *Detection and FeatureExtraction of Atrial Tachyarrhythmias A Three Stage Method of Time-Frequency Analysis BY MARTIN STRIDH,ANDREAS BOLLMANN,S.BERTIL OLSSON,AND LEIF SÖRNMO*.

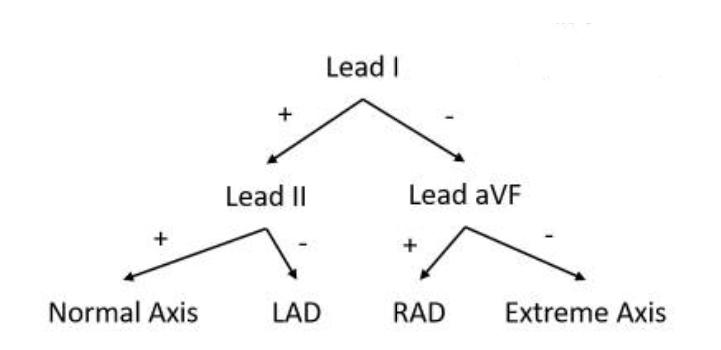
On the boxplots, there is no visual difference. However, the combination between this feature and others could help differentiate between AF and AFL: 12 new features.

(Note 1: I have asked for the code of the above paper)

(Note 2: The paper sent by Leif is performing differentiation between Atrial Flutter and Atrial Tachycardia, not between AFL and AF [**utiliser ces features pour AFL?**]).

**2. Features I will extract for Axis Deviation: LAD/RAD [errors: Misclassification with SNR]:** In electrocardiography, left axis deviation (LAD) is a condition wherein the mean electrical axis of ventricular contraction of the heart lies in a frontal plane direction between −30° and −90°. This is reflected by a QRS complex positive in lead I and negative in leads aVF and II. (Source: [wikipedia](https://en.wikipedia.org/wiki/Left_axis_deviation#:~:text=In%20electrocardiography%2C%20left%20axis%20deviation,in%20leads%20aVF%20and%20II.)).

The condition of LAD is usually defined by a QRS electrical axis and an age (the sane QRS electrical axis varies with the patient’s age).



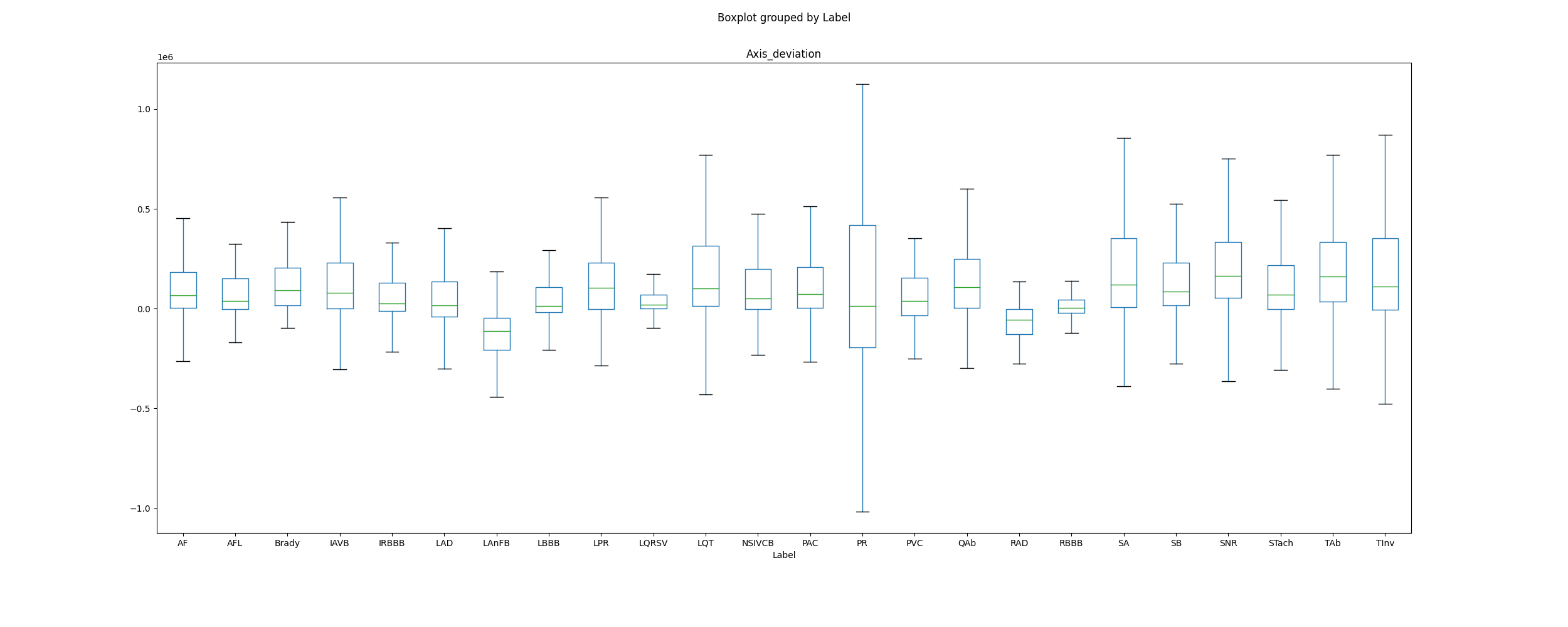
*Method in order to determine the nature of the QRS axis (source:* [*ncbi*](https://www.ncbi.nlm.nih.gov/books/NBK470532/)*):*

In this decision tree, the sign is the sign of the **net QRS deflection**.

The net QRS deflection is computed as follows:

*Note: Controversy when determining the QRS axis with a Bundle Branch Block condition (source:* [*ncbi*](https://www.ncbi.nlm.nih.gov/books/NBK470532/)*).*

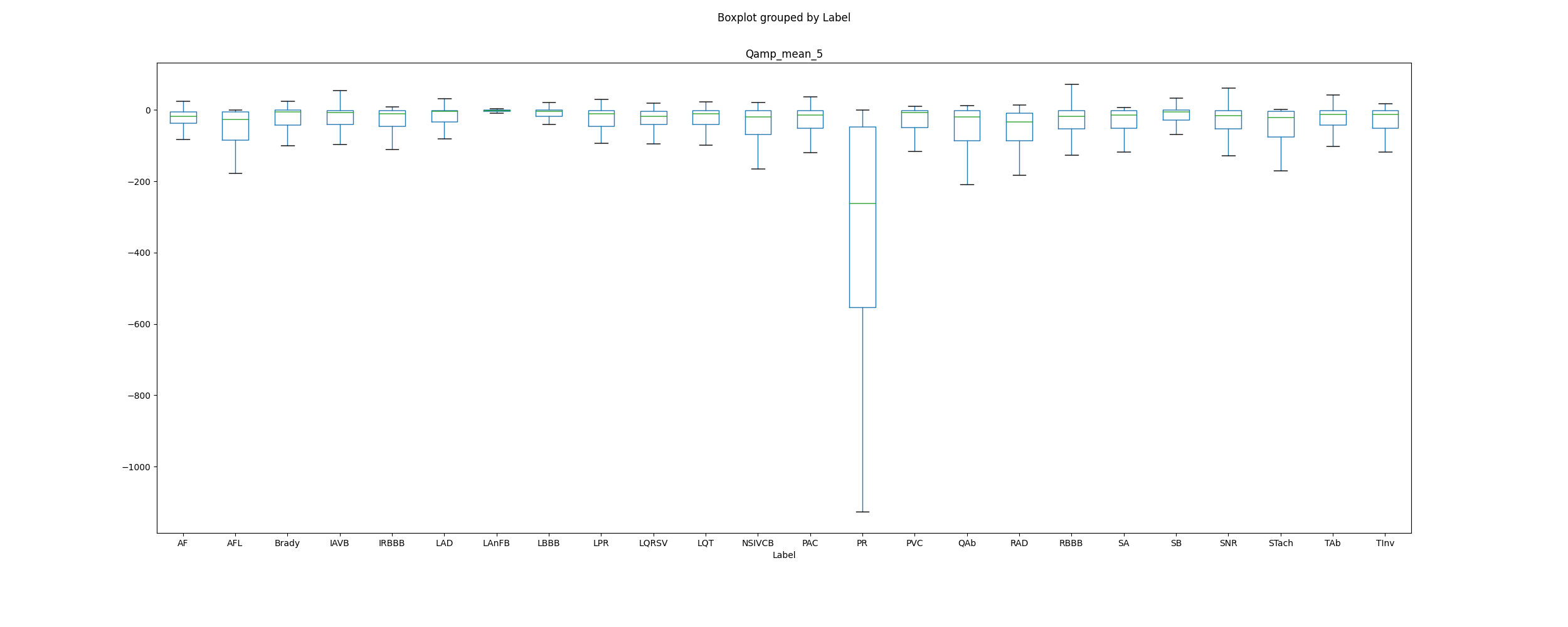
We are going to extract these features for the classification of LAD and RAD. If our classification method is robust, we should perform relatively well because this decision tree structure particularly fits our Random Forest Classifier. The features I will extract for my Classifier are the value of net QRS deflection on leads I, II and aVF and their signs: 6 new features.

Here is the boxplot of: (net\_QRS\_deflection\_lead\_I)\* (net\_QRS\_deflection\_lead\_X) where X = II if the first sign is positive and X = aVF else. Therefore, we should have a negative sign for this feature for the conditions LAD and RAD, with different causes for LAD and RAD.

We can clearly see that this new feature helps to differentiate LAD, LAnFB and RAD from the other pathologies (LAnFB and LAD are the two lowest boxplots): it will help our classification. The values of the LAnFB are differentiated in the boxplot because an ECG characteristics of LAnFB is LAD. Implementing these new features will help us to stop misclassifying LAD/RAD as SNR.

**3. Features I will extract for TAb and TInv [errors: misclassification with AF, LAD and SNR]:** We will extract the frequency of T wave typology among every lead. The wavedet-3D classifies every lead’s T-waves among {normal, inverted, upwards, downwards, biphasic}. Therefore, it accounts for the majority of abnormalities (except ‘Camel-Hump T waves’) and also for the T-wave inversion. Therefore, the frequency of every type of morphology in every lead should help to classify the example as if it presents a Twave abnormality (in the broad sense, also accounting for TInv) or not.

**4. Features I will extract for QAb [errors: misclassification with LAD and SNR]:** The Q-waves are pathologic if they are abnormally wide (>0.2 second) or abnormally deep (>5 mm). ([sciencedirect](https://www.sciencedirect.com/topics/medicine-and-dentistry/q-wave#:~:text=Electrocardiogram%20Interpretation&text=Q%20waves%20represent%20the%20initial,often%20indicators%20of%20ventricular%20hypertrophy.)). Therefore, I will extract features for Q wave duration and Q wave amplitude on every lead. The Q wave amplitude is the difference between the ecg amplitude on Q points and isoelectric lines. The duration of Q wave is the duration between the isoelectric point (QRSon) and the first point such as ecg[]=ecg[].

Therefore, I will extract 5 statistical features for the amplitudes of Q waves and Q waves durations: 10 new features on every lead.

As we can see, these new features will help us to differentiate QAb. Moreover, luckily for us, these features will help us differentiate Pacing Rhythm from other pathologies (why? I do not know 😊).

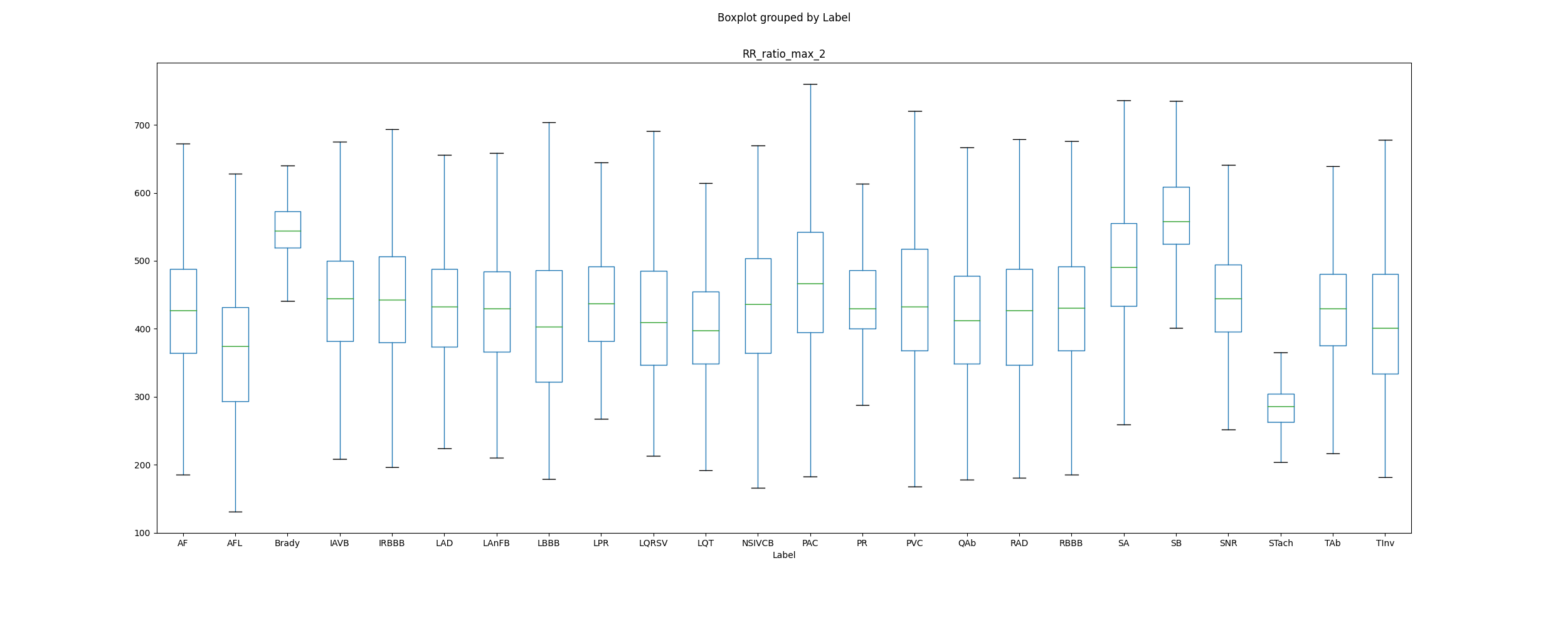
**Qwave duration? Comment tu la calcules? Carré**

**5. Features I will extract for LAnFB (Left Anterior Fascicular Block) [errors: misclassification with SNR]:** ECG characteristics of left anterior fascicular block include the following: QRS slight widening, left axis deviation, small Q and tall R in leads I and aVL (small Q not essential), deep S wave in leads II, III, and aVF (exceeding the R wave) [sources: [sciencedirect](https://www.sciencedirect.com/topics/veterinary-science-and-veterinary-medicine/left-anterior-fascicular-block#:~:text=ECG%20characteristics%20of%20left%20anterior%20fascicular%20block%20include%20the%20following%3A&text=QRS%20width%20normal&text=Left%20axis%20deviation%20(dog%20%3C%20%2B,degrees%2C%20cat%20%3C%200%20degrees)&text=Small%20Q%20and%20tall%20R,aVL%20(small%20Q%20not%20essential)), [ecglibrary](https://litfl.com/left-anterior-fascicular-block-lafb-ecg-library/)]. These are some features I will be extracting for the new pathologies to be studied. I will only add the amplitude of the S\_wave: 5 new features on every lead. Individual boxplot will not be meaningful in this context (although the boxplot for Axis Deviation already pointed out a distinguishable pattern for LAnFB).

**6. Features I will extract for NSIVCB (NonSpecific IntraVentricular Conduction disorder) [errors: Misclassification with SNR]:** Intraventricular conduction disorders are a group of conduction disturbances characterized by abnormalities in intraventricular conduction that leads to changes in shape, duration, and/or axis of the QRS complex on the electrocardiogram. Nonspecific intraventricular conduction disorder exists if the ECG displays a widened QRS appearance that is neither a LBBB nor a RBBB. According to the American Heart Association/ American College of Cardiology and the Heart Rhtyhm society recommendations, non specific intraventricular conduction disorder is defined by a ‘QRS duration greater than 110ms in adults, greater than 90ms in children without meeting the criteria of LBBB and RBBB’. Therefore, the features I have been extracting for PVC and PAC regarding QRS duration and Area should be efficient here. The boxplot did not show any visual evidence of a potential discrimination between pathologies based on QRS duration. Therefore, I do not need to implement additional features for NSIVCB. A possible avenue of amelioration could be seeking for Data Augmentation. (hard to implement new features since the definition of NSIVCB seems not really *strong*).

**7. Features I will extract for SA (Sinus Arrhytmia) [errors: Misclassification with SNR]:**

The ECG criteria to diagnose sinus arrhythmia is a variation of the R-R interval, from one beat to the next, of at least 0.12 seconds, or 120 milliseconds. Therefore, the maximum of the RR interval should help us to see whether an example present the Sinus Arrhytmia or not. Let us plot the boxplot of this feature. (**co occurences: IRBBB**)

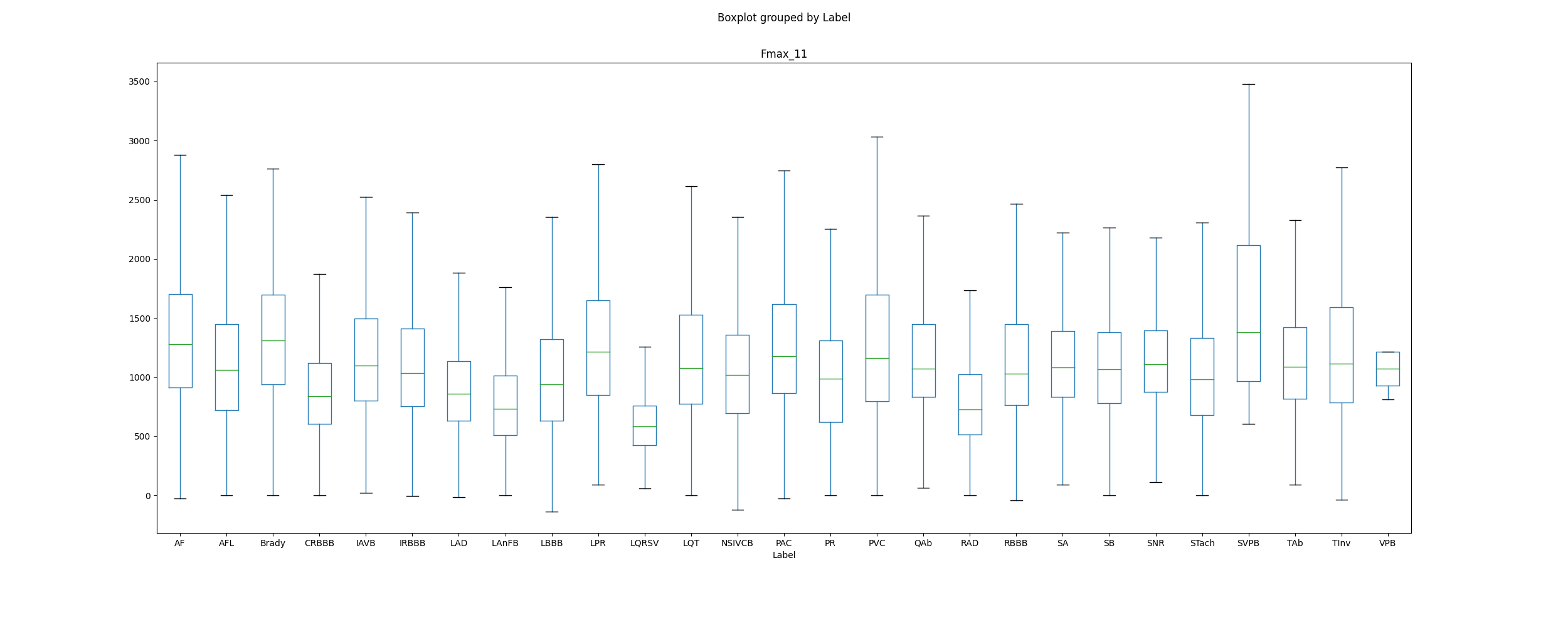


Therefore, we can see that the features of maximal duration between R peaks allow us to separate Sinus Arrhytmias from other pathologies. Therefore, the next step for identifying Sinus Arrhytmia is Data Augmentation.

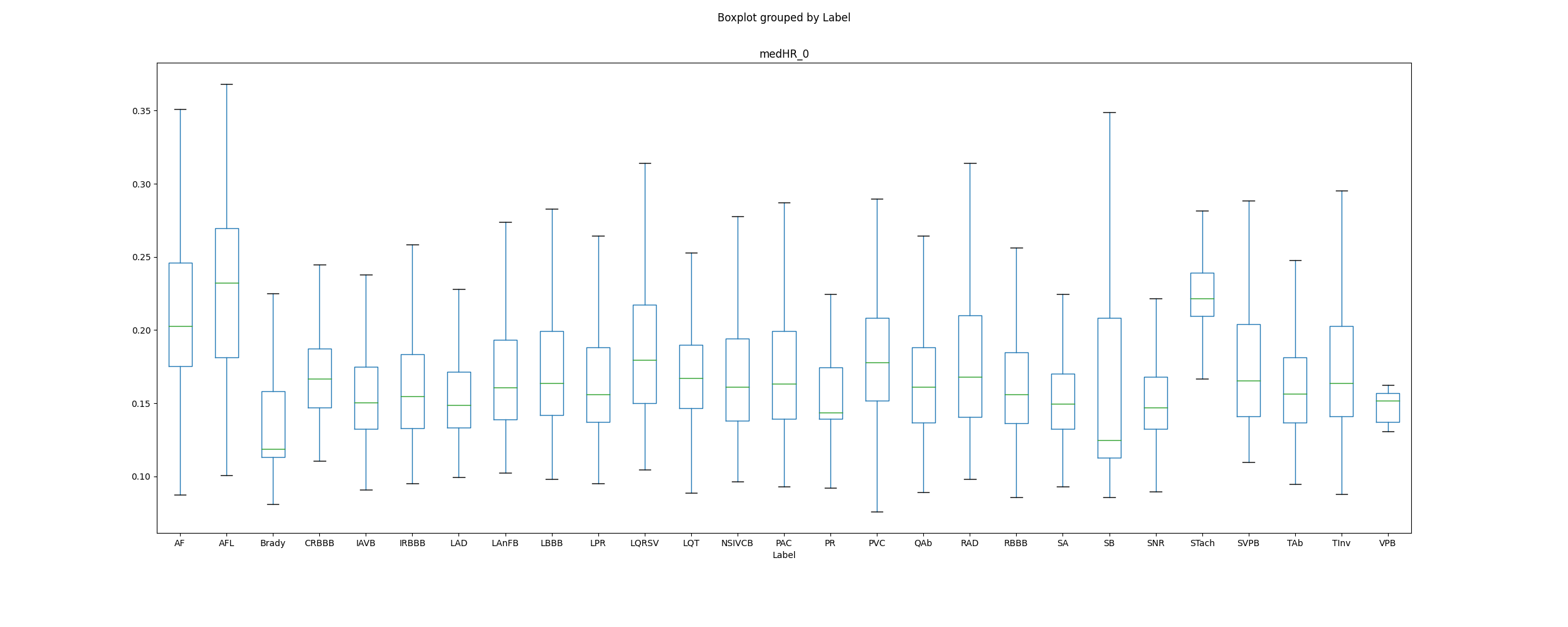
**8. Features I will extract for IRBBB (Incomplete Right Bundle Branch Block) [errors: Misclassification with SNR]:**

Cf Slack conversation avec Joachim

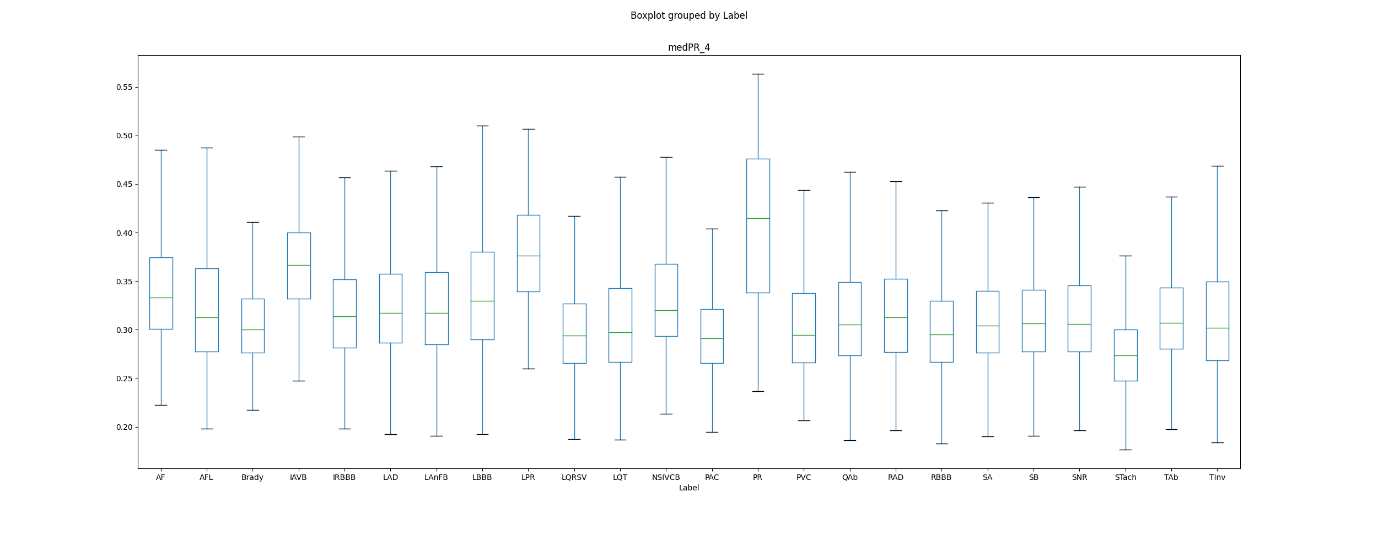
**Pathologies for which we extract relevant features but don’t have enough examples:**

**1. LQRSV [errors: Misclassification with SNR]:** The low qrs voltages can be detected by extracting the amplitudes of R peaks in the ecg.

We can clearly see that this feature enables us to differentiate LQRSV examples from other pathologies: we need to do data augmentation in order to prevent the misclassification with SNR examples.

**2. Brady [errors: Misclassification with SNR]:** The first thing we need to take into account with Bradychardia is that it only co-occurs with PAC, RBBB, IRBBB, PVC. Bradycardia is a condition where an individual has a resting heart rate of less than 60 beats per minute (BPM) in adults. Therefore, the median Heart Rhythm should be a good indicator of Bradychardia.

As we can see from the above boxplot, the median Heart rhythm allows to efficiently separate Bradychardia from the other pathologies. The main source of error from Brady comes from misclassification with SNR, though this feature should help to separate between both. Therefore, we misclassify as SNR out of bias towards the most represented class. This is why the next step for Brady is Data Augmentation.

**3. LPR [error: misclassification with SNR]:**

Just with this feature, our model should be able to differentiate between LPR examples and SNR examples: we need to augment the data we have.

**4. PVC [error: misclassification with SNR]:** I know that the several features I extracted for PVC are discriminative between PVC and SNR. The cause of the low scores are therefore the lack of examples for our Classifier. [**travailler au Beat par Beat**]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Pathology** | **Pathologies misclassified with** | **Additional Features** | **Data Augmentation (immediate next step, supposes we have solved the weighted class problem)** | **Data Augmentation immediate availability** |
| AFL | AF | Max\_freq | Yes |  |
| LAD/RAD | SNR | Net QRS deflections | No |  |
| TAb/ TInv | AF, LAD, SNR | Frequence of Twave typology | No |  |
| QAb | AF, SNR | Depth and Duration of Qwave | No |  |
| LAnFB | SNR | No | No |  |
| NSIVCB | SNR | No | No |  |
| SA | SNR | No | No |  |
| IRBBB | SNR |  | No |  |
| LQRSV | SNR | No | Yes |  |
| Brady | SNR | No | Yes |  |
| LPR | SNR | No | Yes |  |
| PVC | SNR | NO | Yes |  |

**Pathologies for which I do not understand the low results:** PAC, IAVB.

I do not understand how the scores of PAC have changed from 0.72 to 0.0 when we started performing Multi-Label Classification. I do not understand also why the score of IAVB is that low though we extract some high quality features.

**Next Steps:**

**The first global Next Step will be to find out how to perform Hyperparameter Tuning on a OneVsRest Classifier (this will allow to tune our score and perform weighted Classification).**

**I will also extract the features I listed**

**I will also extract these features on new databases in order to augment our training DataBAse**

**Time Expected in order to conduct these experiments:** One week.

**A faire :**

Regarder: deux manières de faire les choses :

OneVsRest Classifier for multitask classification

Deuxième option: 1 seul Classifier mais je factorise le loss pour chacun des labels (slides Joachim, dernier slide)

LQRSV : delta entre le point R et le point S (trouver une définition plus claire) (site de Joachim).

Data Augmentation : mettre plus de weight que les autres examples (ce qui n’est pas le cas des bases du Challenge)

PAC : regarder la raw data et les exemples 1 par 1 pour voir ce qu’il se passe

Prefiltering important

Feature Selection and Class imbalanced

MultiLabel alternative method (pas sûr que ça marche mieux mais sdv)

Hyperparameter tuning

Extract new features

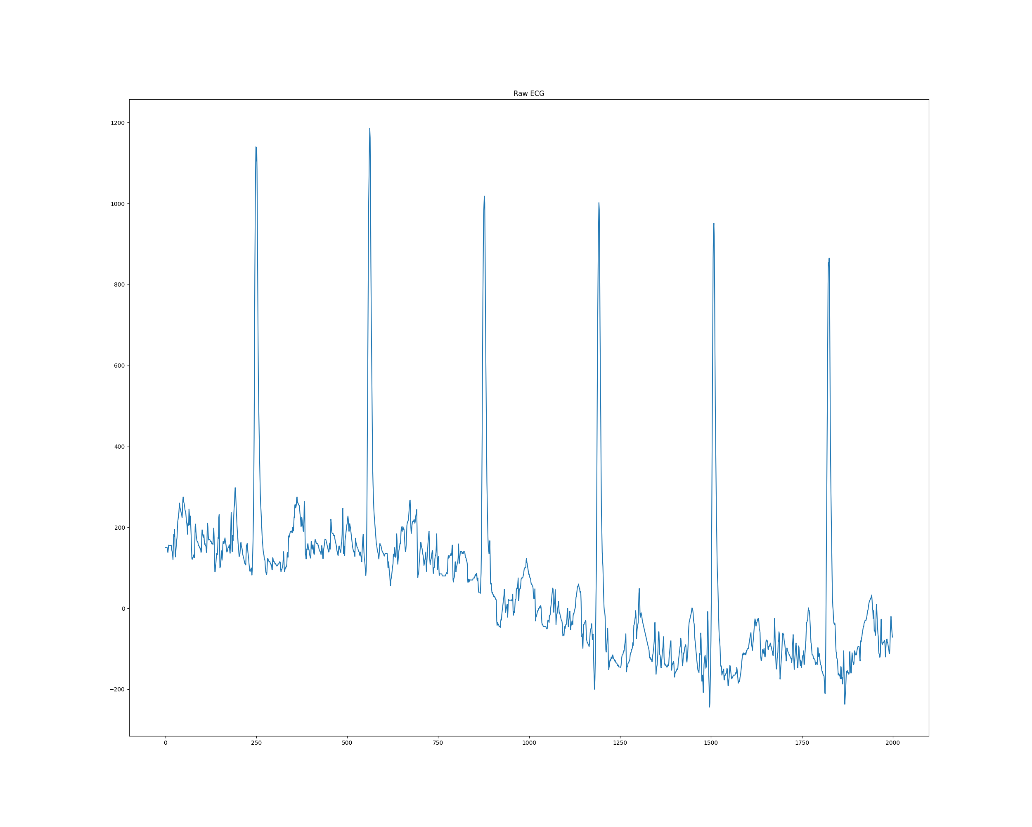
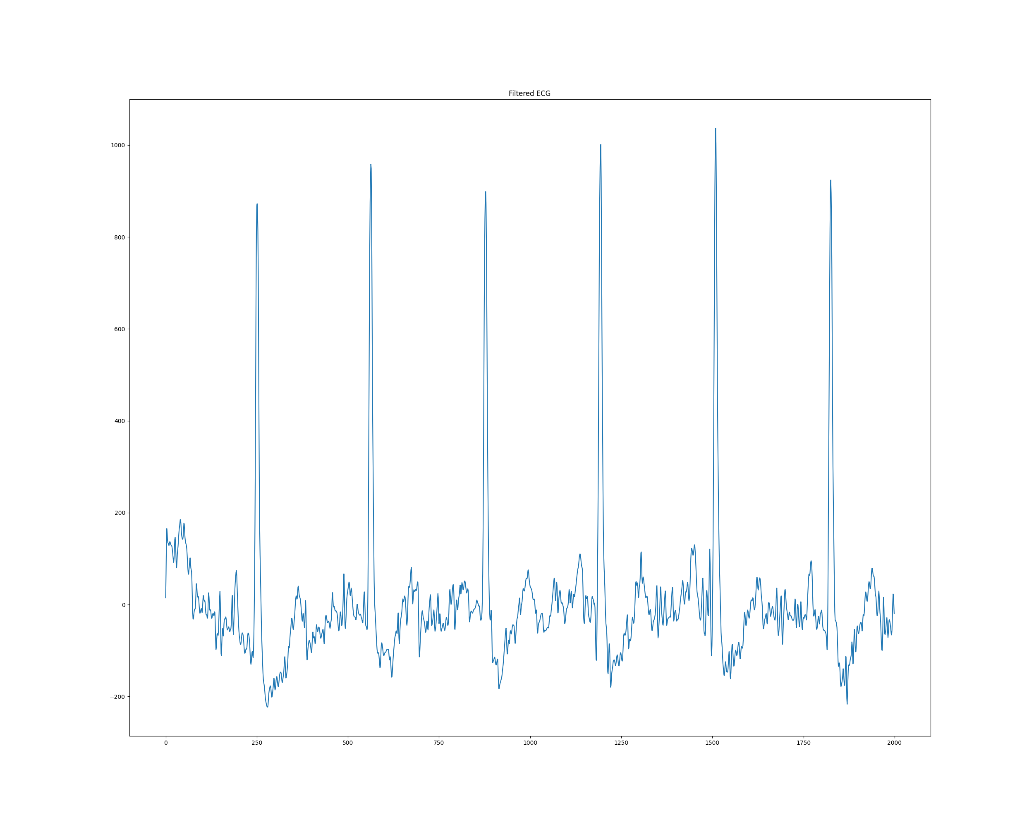
Data Augmentation with new databases

**PreFiltering step:**

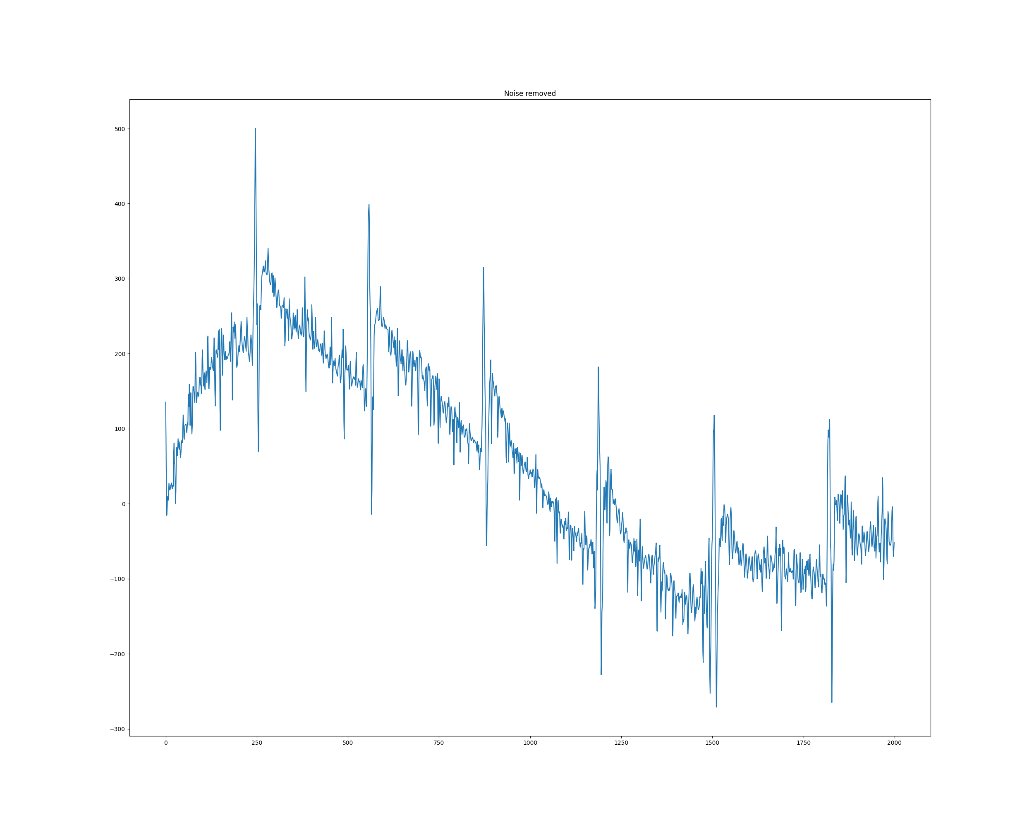
**Noises to be removed: Baseline wandering, Tremor noise, Power Line Interference and HF Noise.**

The filter we used: A bandpass filter [0.05-100] and I have added two notch filters [58;62] and [48;52] in order to remove power line interferences (coming from Europe or Asia?)

**Visualisation of the filtering step**

****

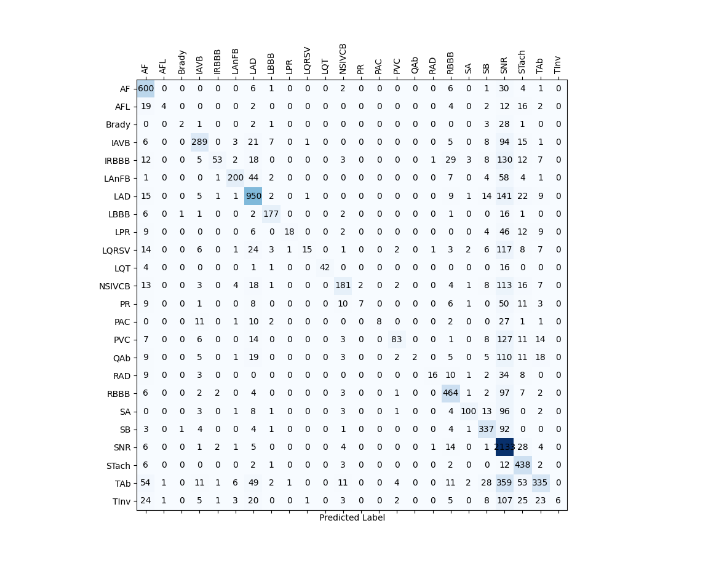
Raw ECG Filtered ECG

****

Noise removed

**End of Feature Extraction with preprocessed signal: Saturday Night: Classifier Results on Sunday**

**Class Weighting Tentative:**

Failure in Hyperparameter Tuning for now but I tried to set the weighting parameters on the default estimator. Here is the confusion matrix of the resulting model: **no conclusive.**

Next avenues to try: resampling, Data Augmentation

**Next steps:**

Feature Selection (remote computer): end by **Thursday**

Data Augmentation (local computer): end by **Wednesday**

Model submission: end by **Wednesday**

Resampling if necessary: end of next week

Try other multi-label strategies: end of next week

Extract other features: next week-end

HyperParameter tuning on the final RF/XGBoost Model: next week end