|  |  |  |  |
| --- | --- | --- | --- |
| **Pathologies** | **Number of Examples** | **Number of Features extracted specifically for this pathology per lead** | **Fbeta score** |
| AF | 2345 | 16 | 0.84 |
| AFL | 308 | 16 (the ones from AF)[[1]](#footnote-1) | 0.25 |
| Brady | 259 | 0 | 0.26 |
| IAVB | 1318 | 3[[2]](#footnote-2) | 0.61 |
| IRBBB | 1221 | 0 | 0.02 |
| LAnFB | 1254 | 0 | 0.42 |
| LAD | 2126 | 0 | 0.57 |
| LBBB | 982 | 20[[3]](#footnote-3) | 0.85 |
| LPR | 338 | 6 (3 AVB, 3 challenge) | 0.11 |
| LQRSV | 526 | 0 | 0.15 |
| LQT | 1090 | 5 (challenge) | 0.33 |
| NSIVCB | 897 | 0 | 0.04 |
| PR | 299 | 0 | 0.51 |
| PAC | 1337 | 12[[4]](#footnote-4) | 0.76 |
| PVC | 552 | 26[[5]](#footnote-5) | 0.18 |
| QAb | 824 | 0 | 0.01 |
| RAD | 403 | 0 | 0.22 |
| RBBB | 2018 | 20 (the ones from LBBB) | 0.70 |
| SA | 1087 | 0 | 0.50 |
| SB | 1606 | 0 | 0.64 |
| SNR | 12019 | 0 | 0.80 |
| STach | 1555 | 0 | 0.82 |
| TAb | 1865 | 0 | 0.14 |
| Tinv | 832 | 0 | 0.03 |

**Pathology Analysis**

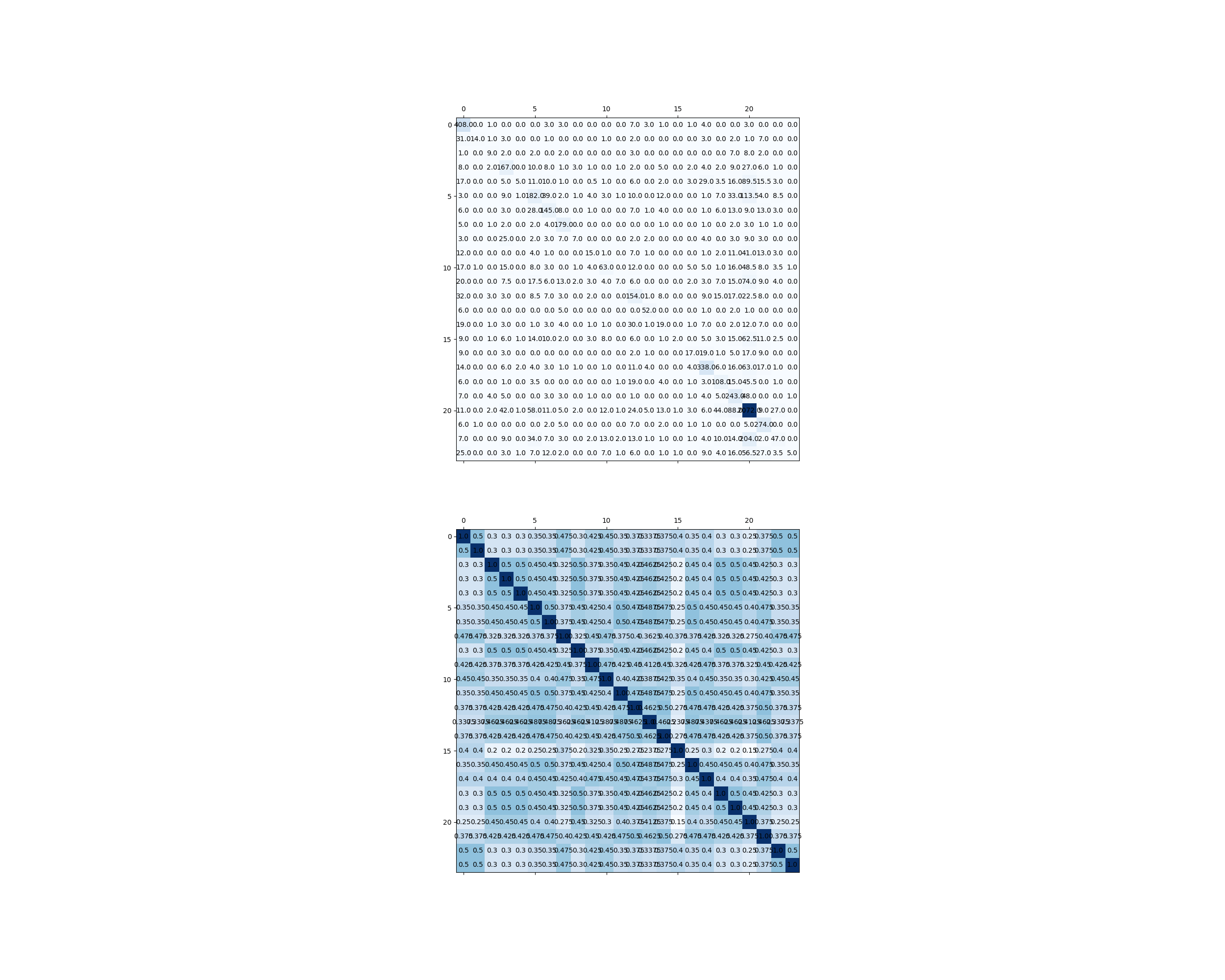
**Things that we need to understand**

**Three Cases:**

**1. Features but not enough examples:** We need to find new DataBases in order to augment our existing ones and see how these pathologies behave with others.

Pathologies that are concerned: AFL (available examples to augment with Independent DataBase 1 and 2: 500), PVC (available examples to augment with Independent DataBase 1 and 2: X), LPR (available examples to augment with Independent DataBase 1 and 2: X)

**2. Enough Features, Enough examples and low scores:** We need to understand why do we perform poorly with such pathologies. Let us recall the confusion matrix.

****Pathologies concerned: IAVB (label number 3), LQT (label number 10), PVC (label number 14).

**For IAVB:** The serious mistake we make are confusing AVB with SNR

**For LQT:** We mistake LQT for AF, AVB and SNR

**For PVC:** We mistake PVC for AF, LPR and SNR.

**Question:** How can we prevent examples from being misclassified as SNR?

**3. Enough examples, not enough Features:** We need to find new features to extract for these pathologies. Pathologies concerned: Brady, IRBBB, LAnFB, LAD, LQRSV, NSIVCB, PR, QAb, RAD, SA, SB, TAb, TInv, LPR

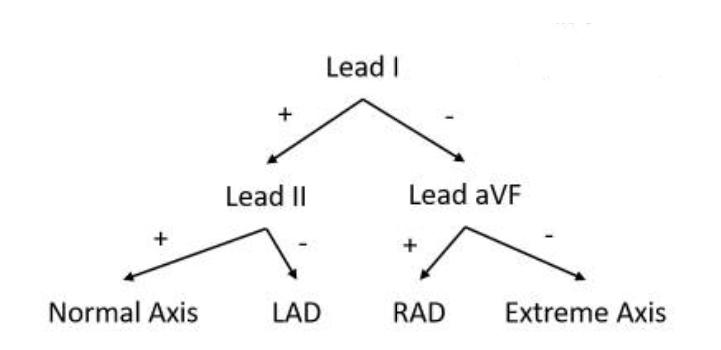
**Let’s start with the morphological pathologies:**

**A. LAD/RAD: Axis Deviation**

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/)*).*

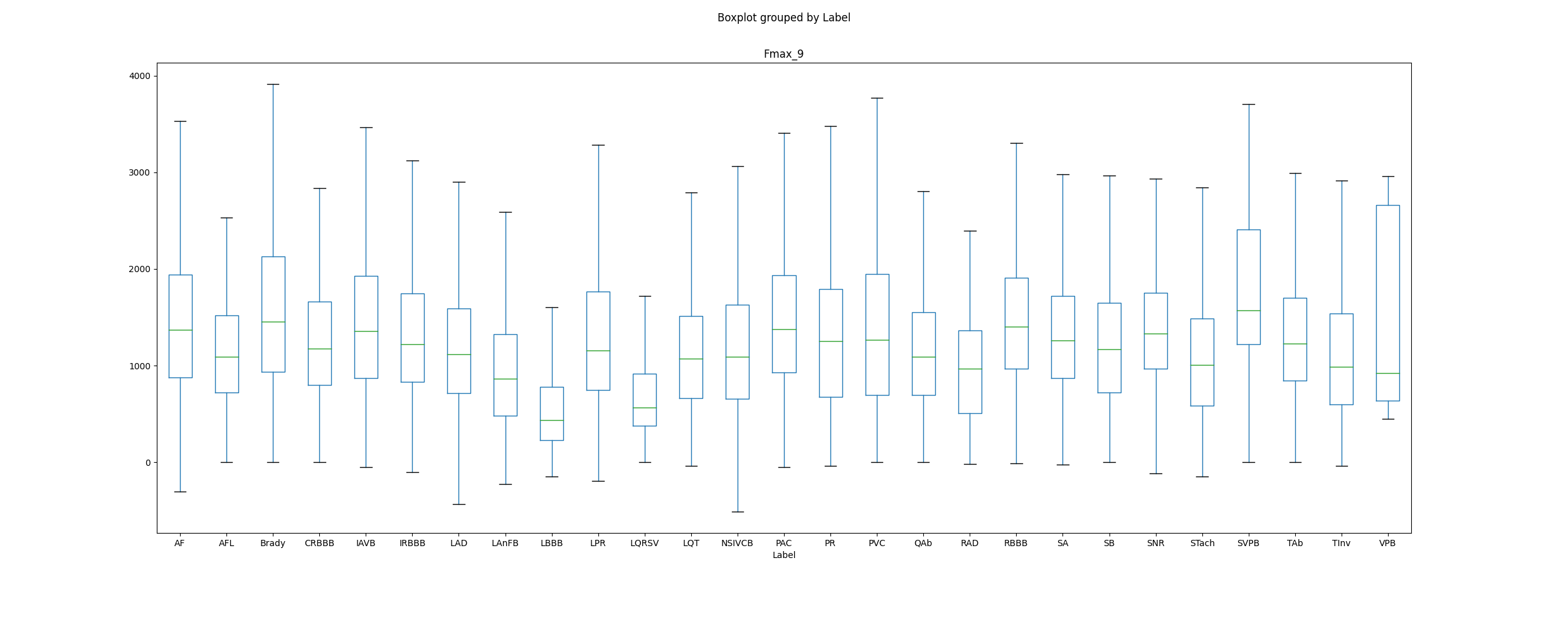
We are going to extract these features for the classification of LAD and RAD. If our classification method is robust, we should perform relatively well because this decision tree structure particularly fits our Random Forest Classifier. The features I will extract for my Classifier are the value of net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

Here is the boxplot of: sign(net\_QRS\_deflection\_lead\_I)\*sign(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.

*Note: The boxplot does not take into account multi-labelled examples, thus it may explain why some other pathologies present the characteristics of Axis Deviation (secondary pathologies of these examples may be RAD/LAD)*

**B. LQRSV**

Low voltage on the ECG is defined as a peak-to-peak QRS amplitude of less than 0.5 mV in the limb leads (I, II and III) and/or less than 1 mV millimeters in the precordial leads (V1 to V6). [*Low QRS voltage and its causes, John E.Madias*]

However, I already extract statistical features of QRS amplitudes (maximum, std, mean, median): why do we perform poorly on this pathology? (I was not extracting the minimum of amplitudes, I will from now on).Boxplots of leads I, II, III and V1 to V6 of QRS amplitude statistical features:

Therefore, we can see that some behaviors seem specific to LQRSV: low amplitudes and that this behaviour is captured by some features. Therefore, the cause of low score for LQRSV: **not enough examples.**

**C. QAb**

They 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[].

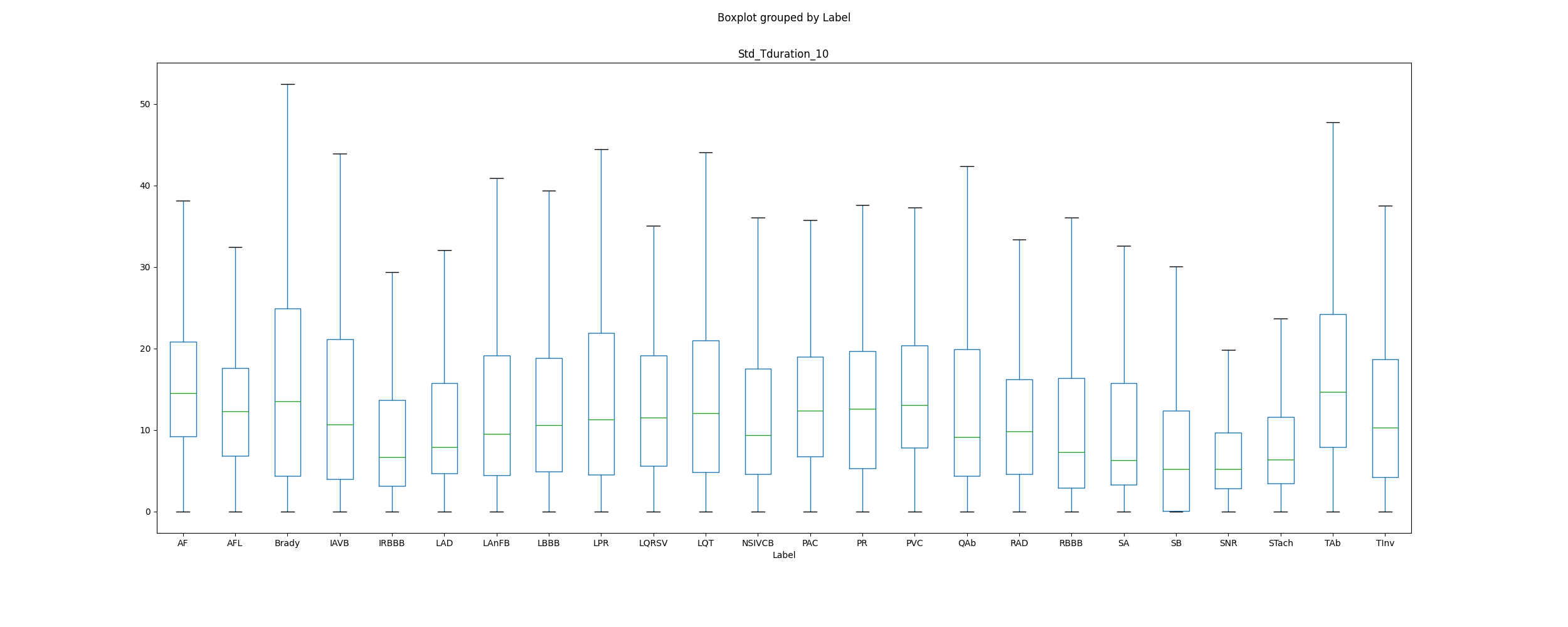
Therefore, I will extract 5 statistical features for the amplitudes of Q waves and Q waves durations: 10 new features on every lead.

*Note: The features I have extracted are not discriminative for QAb (boxplots non conclusive but are for very discriminative for pacing rhythm: I will keep them and look for new ones for discriminate QAb.)*

**D. TAb/TInv: Information about the Twave**

We will extract the following features: statistical features for the amplitude (flattened and Inverted T waves) and duration of the T wave (peaked a,d Hyperaccute T wave), T typology (from the wavedet\_3D output: in {normal, inverted, upwards, downwards, biphasic}), T typology on and T typology off: 25 new features on every lead.

Missing type of Abnormality: ‘Camel-Hump T waves’

The boxplot of new features extracted:

The typologies features are not interesting since using statistical quantities destroy valuable information.

1. I have tried the f\_wave detection (with the frequency of the maximum of the ecg, this did not successfully separated AF from AFL) [↑](#footnote-ref-1)
2. 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 [↑](#footnote-ref-2)
3. 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 [↑](#footnote-ref-3)
4. 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 [↑](#footnote-ref-4)
5. 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 during the unofficial phase [↑](#footnote-ref-5)