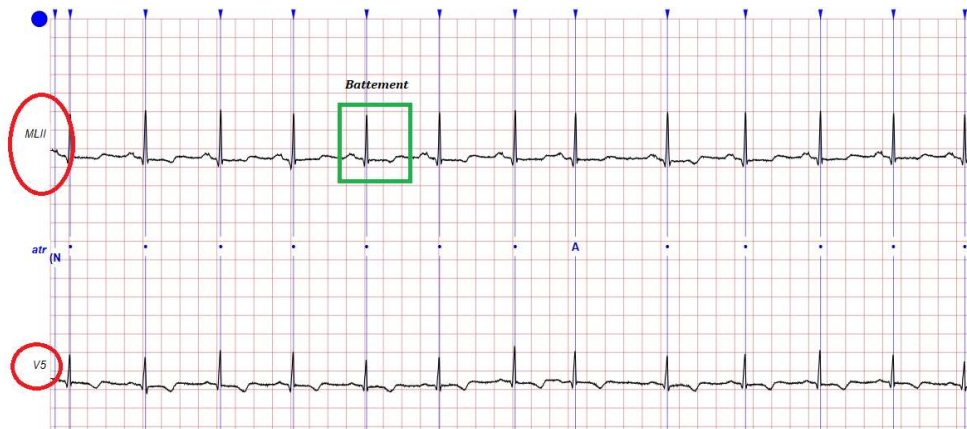


Descriptif de la base MIT-BIH Arrhythmia Database :

La base contient 48 extraits de 30 minutes chacun, enregistrés chez 47 patients. Chaque enregistrement ECG comprend deux canaux (ou dérivation) :

- MLII (Modified Lead II) : c'est la dérivation principale, souvent utilisée pour l'analyse des battements cardiaques car elle offre une meilleure lisibilité des ondes P, QRS et T. **Nous allons travailler avec ce canal, car il est plus représentatif du rythme cardiaque.**
- V1 (ou autre dérivation précordiale V2->V5) : utilisée comme complément pour une analyse plus complète.



Chaque enregistrement contient 4 fichiers :

- **Header file** (*.hea): contient les métadonnées (nombre de canaux, fréquence d'échantillonnage, durée, noms des dérivation, etc.).
- **Signal file** (*.dat): fichier binaire contenant les données ECG brutes, encodées selon les infos du header.
- **Annotation file** (*.atr) : positions des battements et leurs types (normaux, ventriculaires, ectopiques, etc.). Très utile pour créer les labels.
- **Fichier optionnel** (*.xws): utilisé pour l'affichage graphique dans certains outils. Non indispensable pour notre analyse ML.

Sur un tracé ECG, chaque battement est représenté par un ensemble d'ondes caractéristiques, appelées le complexe PQRST. Ces battements sont détectés par le pic R, qui est le point le plus haut du complexe QRS. Les fichiers .atr contiennent la position exacte de chaque pic R (timestamp) ainsi que le type de battement associé (annotation médicale : normal, ventriculaire, etc.).

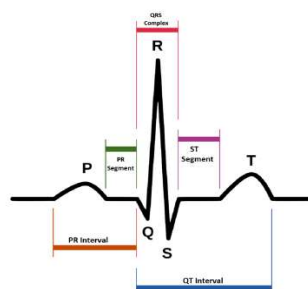


Figure 1. Morphology of a normal ECG.

Un enregistrement complet dure 30 minutes, ce qui représente des centaines de battements. Il n'est pas nécessaire d'utiliser tout l'enregistrement en une seule fois pour classer un rythme. La plupart des anomalies peuvent être identifiées en regardant la forme locale d'un seul battement.

Dans ce projet, nous allons adopter une approche de Classification par battement (Beat-wise classification). Cette méthode consiste à extraire de courtes fenêtres centrées sur chaque R-peak, par exemple 0.2 secondes avant et 0.4 secondes après le pic, puis classer chaque battement individuellement, en utilisant les annotations fournies comme labels d'apprentissage. Voici donc le pipeline que vous devez suivre :

1. Détection des R-peaks (présents dans les .atr)
2. Extraction d'un segment court (ex: 200 ms avant et 400 ms après le R)
3. Prétraitement : normalisation, resampling, etc.
4. Appliquer un modèle ML ou DL et classification du battement