**ECG Documentazione**

**Dati di physionet e kaggle**

Sono stati utilizzati gli ECG del dataset di physionet (MIT-BIH Arrhythmia Database <https://www.physionet.org/content/mitdb/1.0.0/),la>

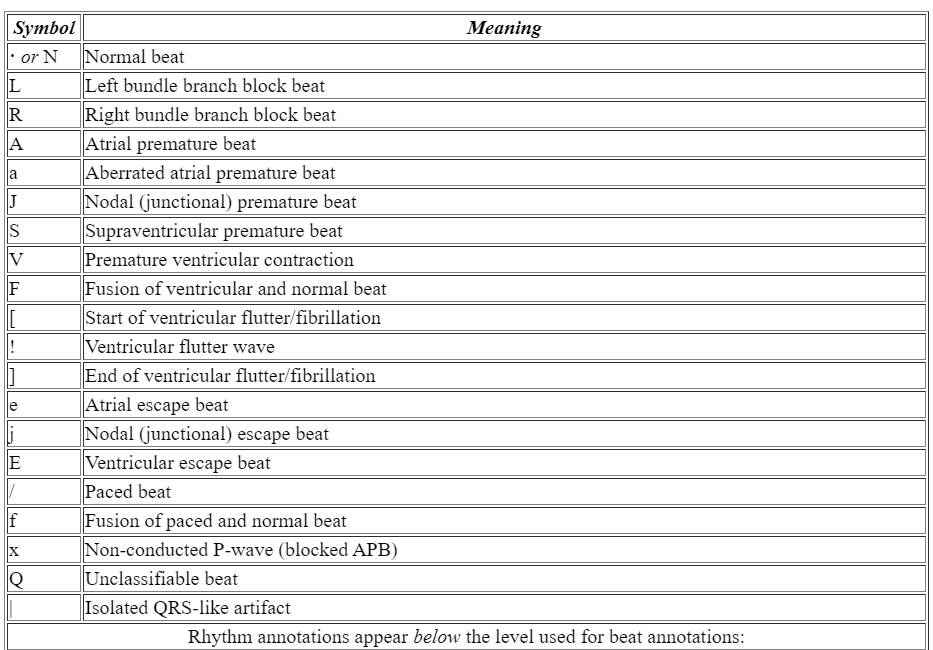
Documentazione del dataset : <https://archive.physionet.org/physiobank/database/html/mitdbdir/mitdbdir.htm>

Il database delle aritmie del MIT-BIH contiene 48 estratti di mezz'ora di registrazioni di ECG ambulatoriali a due canali, ottenuti da 47 soggetti studiati dal Laboratorio di Aritmia della BIH tra il 1975 e il 1979. Ventitré registrazioni sono state scelte a caso da un set di 4000 registrazioni di ECG ambulatoriali a 24 ore raccolte da una popolazione mista di pazienti ricoverati (circa il 60%) e ambulatoriali (circa il 40%) presso il Beth Israel Hospital di Boston; le restanti 25 registrazioni sono state selezionate dallo stesso set per includere aritmie meno comuni ma clinicamente significative che non sarebbero ben rappresentate in un piccolo campione casuale.

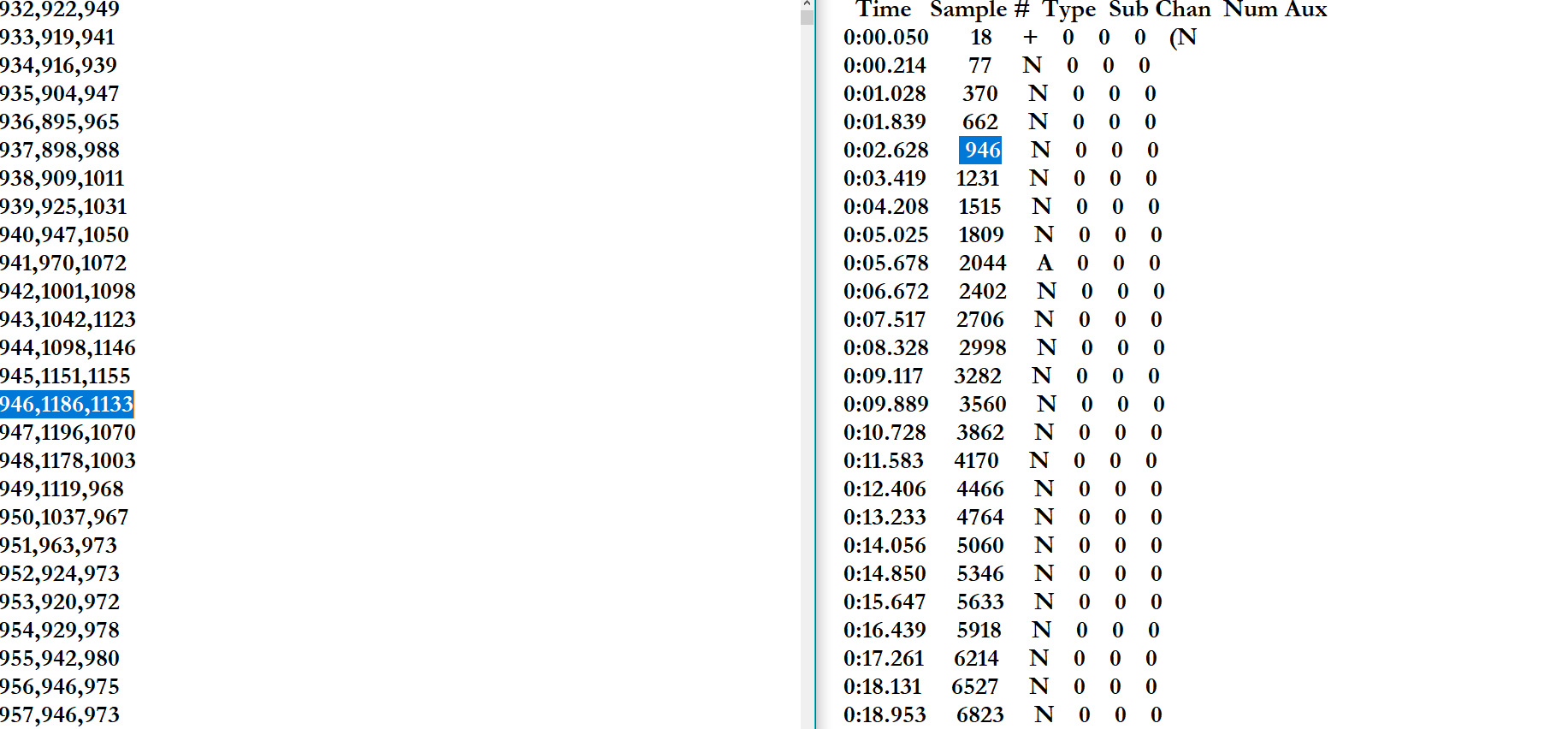
Le registrazioni sono state digitalizzate a 360 campioni al secondo per canale con risoluzione a 11 bit su un range di 10 mV. Due o più cardiologi hanno annotato indipendentemente ogni registrazione; i disaccordi sono stati risolti per ottenere le annotazioni di riferimento leggibili dal computer per ogni battito (circa 110.000 annotazioni in tutto) incluse nel database.

Da kaggle <https://www.kaggle.com/taejoongyoon/mitbit-arrhythmia-database> sono stati presi i csv che descrivono i valori delle derivazioni dei campioni degli ecg di physionet.

Inoltre sono stati utilizzati i file di annotazione presi sempre da kaggle, in cui sono i vari picchi (complesso QRS) sono etichettati con un simbolo che rappresenta le anomalie trovate.



**Sample,MLII,V5**



**Panoramica sul dataset**

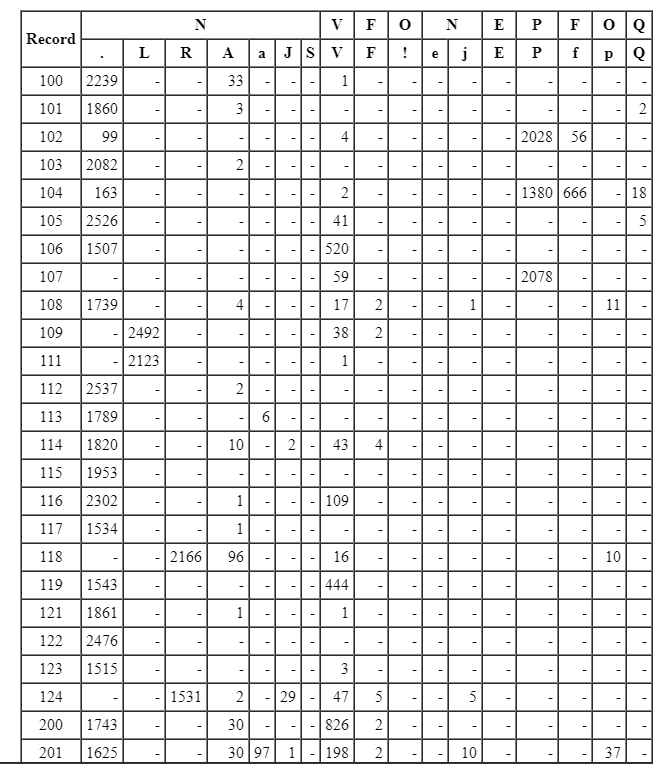
[MLI,V1] = 101,105-113,115,116,118-122,200- 231

[MLI,V5] = 100,114(inverse)

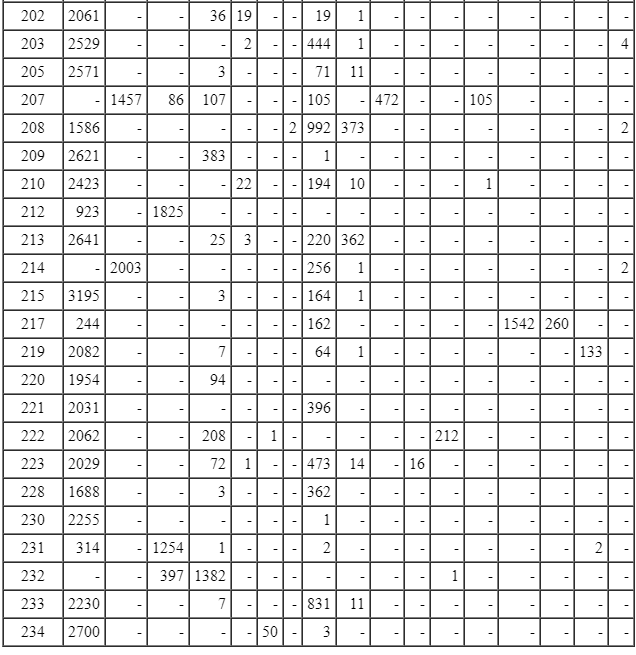
[MLI,V4] = 124

[MLI,V2] = 103,117,123

Tracciati con le relative anomalie rilevate







**Segnali registrati**

Nella maggior parte dei record, il segnale superiore è un cavo modificato dell'arto II (**MLII**), ottenuto posizionando gli elettrodi sul torace.

Il segnale inferiore è di solito un elettrocatetere modificato **V1** (occasionalmente **V2** o **V5**, e in un caso **V4**); come per il segnale superiore, anche gli elettrodi sono posizionati sul torace.

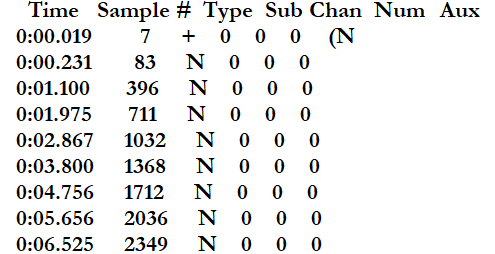
I normali complessi QRS sono di solito prominenti nel segnale superiore (**MLII**)

**Preprocessing del Dataset**

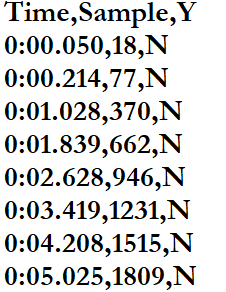
Per poter utilizzare i dati creando un dataset finale, abbiamo fatto le seguenti operazioni:

1. È stato modificato il formato dei files di annotazione(100annotations.txt,101annotations.txt) in modo da poter essere processati ed essere letti come un file csv, rimuovendo alcuni valori superflui.

Prima:



Dopo:



I file sono memorizzati nella cartella “files\_preprocessed”

Il codice di questa procedura è presente in ***preprocess\_annotations.py***

1. Le tracce ECG da 30 minuti (i file 100,101,...) sono state frammentate in segmenti da 30 secondi, quindi sono stati presi per ogni segmento 10800 campioni , si tiene conto solo della derivazione **MLII** per ogni campione del segmento, per questo motivo sono stati scartati i file 102 e 104 in quanto presentavano solamente le derivazioni V1 e V5  
     
   *L'etichettatura di ogni segmento dipende dalle annotazioni assegnate ai vari picchi presenti nel segmento ,quindi ogni segmento di 30 secondi avrà una sola etichetta stabilità nel seguente modo:  
     
    - Se nel segmento di 30 secondi tutti i picchi sono annotati con* ***N (normale)*** *l'intero segmento verrà etichettato con la label N*

* *Se invece ci sono dei picchi etichettati con un simboli diversi da N che rappresentano vari tipi di anomalie, il segmento verrà etichettato con* ***A (Anomalia)***

Per ogni tracciato di 30 minuti è stato creato un file csv ( data100.csv ,data101.csv …) dove ogni riga rappresenta un segmento di 30 secondi etichettato con N o con A.

Per avere un unico csv rappresentate tutto il dataset,tutti i csv dei tracciati prima descritti sono stati uniti in un unico csv, questo file è presente nella cartella dataset\full\_dataset.

Il codice di questa procedura è presente in ***generate\_dataset.py***