**ECG CNN Diagnostic System**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

L'invecchiamento della popolazione sta comportando un incremento dei pazienti che soffrono di patologie cardiache e che quindi necessitano di un monitoraggio elettrocardiografico.

Già oggi la refertazione degli elettrocardiogrammi normali comporta un notevole impiego di tempo da parte degli specialisti cardiologi, la cui preparazione è oltremodo lunga e costosa.

Sarebbe quindi importante poter escludere il maggior numero possibile di ecg che sicuramente non presentano anomalie dalla necessità che siano refertati dallo specialista evitando però nel modo più assoluto che la presenza di una alterazione elettrocardiografica, che potrebbe essere la spia di una patologia cardiologica, non sia riconosciuta e sfugga all'ossrvazione del cardiologo.

Allo stesso tempo si vanno diffondendo numerosi device come Kardia ([www.quiver.store](http://www.quiver.store/)) e dhearth (<https://www.d-heartcare.com/>) che permettono ai cittadini di registrare in modo autonomo il proprio ecg. E' prevedibile che il numero di tracciati registrati nei prossimi anni cresca in modo esponenziale e non è assolutamente pensabile che tutte le registrazioni prodotte potranno essere valutate da un cardiologo.

Devono quindi essere sviluppati dei sistemi che possano in modo automatico selezionare quali sono i tracciati che necessitano di valutazione specialistica e possibilmente anche l'urgenza di tale valutazione.

Per questi motivi è stata proposta un architettura di rete neurale convoluzionata (CNN) per lo sviluppo di un sistema automatico in grado di riconoscere la presenza o meno di un’ anomalia all’interno di un elettrocardiogramma (ECG) .La CNN quindi effettuerà una classificazione binaria degli ECG, Normale (N) o con Anomalie (A).

**Dati utilizzati**

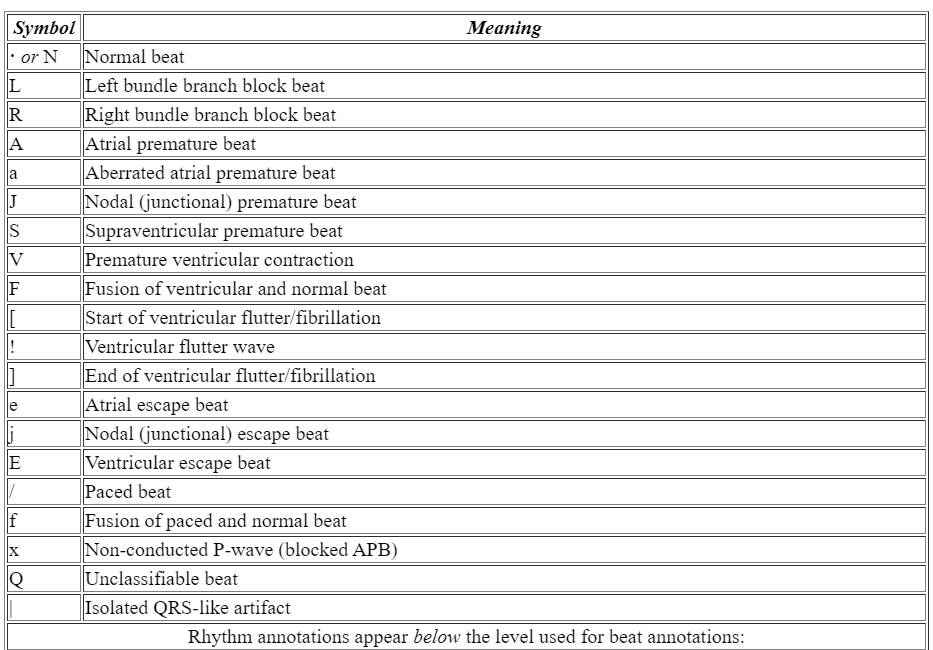
E’ stato utilizzato il dataset MIT-BIH Arrhythmia Database da PhysioNet e da Kaggle.

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. 23 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 una gamma 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.

Kaggle contiene i vari campioni dei tracciati ECG del database MIT-BIH Arrhythmia Database di PhysioNet, strutturati in csv e le annotazioni dei picchi registrati nei tracciati in file di testo.

Le annotazioni relative ai picchi sono le seguenti:



Per ulteriori informazioni sul dataset utilizzato si possono consultare i seguenti link:

-https://archive.physionet.org/physiobank/database/html/mitdbdir/intro.htm#symbols

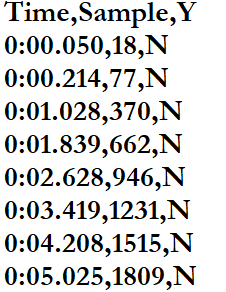
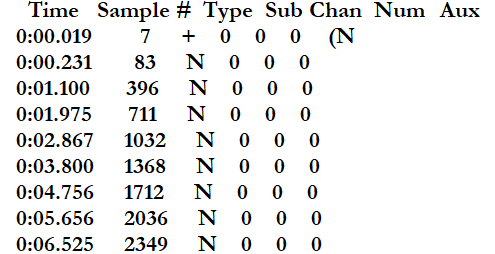
-https://archive.physionet.org/physiobank/annotations.shtml

**Preprocessing del Dataset**

Partendo i file presenti in kaggle, sono state effettuate le seguenti procedure

1. È stato modificato il formato dei files di annotazione eliminando dati superflui in modo da poter essere processati come un file csv.

Esempio



1. I tracciati ECG da 30 minuti sono stati frammentati in segmenti da 30 secondi ciascuno, quindi per ogni segmento si hanno 10800 campioni , si sono memorizzati solo i valori 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
2. In base alle annotazioni assegnate ai vari picchi presenti,ogni segmento è stato etichettato nel modo seguente:  
    - Se nel segmento 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 dove ogni riga rappresenta un segmento 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.

Il dataset risultante presenta **871** esempi di segmenti etichettati come **Normali (N)** e **1769** etichettati come **Anomali (A)**.

**Architettura della CNN**

Il modello utilizzato per la classificazione binaria (tracciato Normale o tracciato con Anomalie) di ECG è una rete convoluzionata sequenziale con 5 layer , con la seguente struttura:

I primi 4 layer sono composti da:

- Conv1D

- BathcNorm1D

- MaxPool1D

- Funzione di attivazione RELU

L’ultimo layer è composto:

- AveragePooling1D

- Layer Flatten

- Layer Dense con funzione di attivazione Softmax

La figura seguente mostra un estratto della struttura della rete neurale implementata in Python utilizzando la libreria Keras.

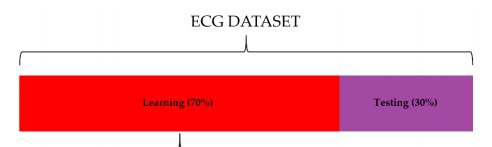
Tale modello è stato realizzato ispirandosi alla rete neurale proposta dall’articolo ‘Automatic ECG Diagnosis Using Convolutional Neural Network’

**Training Validation e Testing**

L’ intero dataset e stato diviso nel seguente modo:

- 70% di **Learning Set**

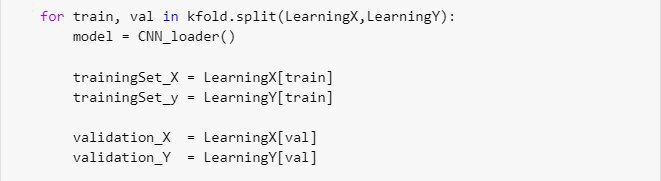
- 30% di **Test Set**



Per la **validazione** degli iperparametri del modello è stata utilizzata la *convalida incrociata K-Fold* con k=10 sul LearningSet.

A ogni iterazione del k-Fold si sono effettuate le seguenti operazoni:

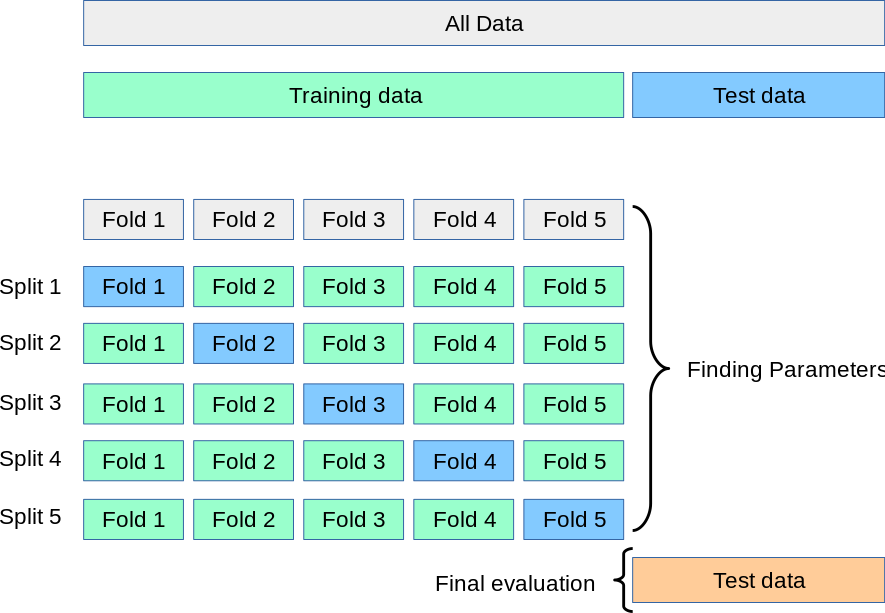
- Creazione della CNN

- Suddivisione del LearningSet in k parti dove k-1 parti sono state utilizzate per l’ addestramento (TrainingSet) e la restante parte come ValidationSet

- Calcolo dell’accuratezza e il recall

Alla fine delle varie iterazioni si è calcolata la media e la deviazione standard delle accuratezze e del recall.

Nella figura seguente è presentato un esempio di convalida incrociata k-fold con k=5



Dopo aver effettuato la **validazione** è stata effettuata la **predizione**.

Per la **predizione** la CNN è stata addestrata sull’ intero LearningSet e successivamente si è effettuata la predizione per i dati del TestSet utilizzando il modello appena creato.

In fine si è visualizzata la matrice di confusione e le seguenti metriche:

- Precision =

- Recall=

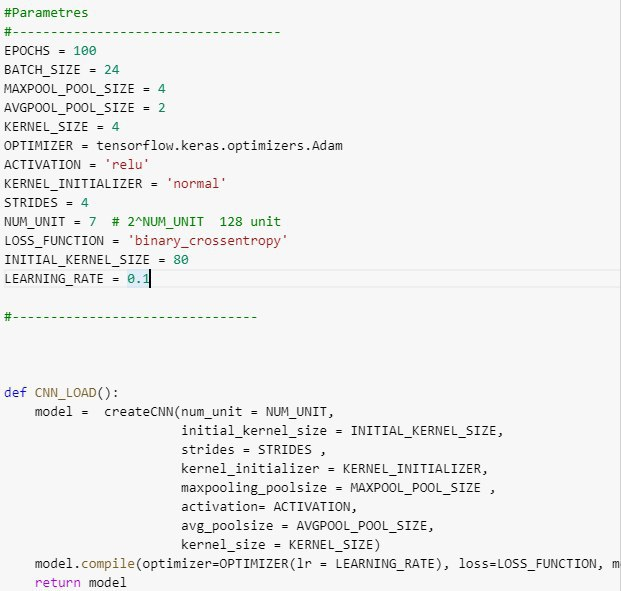
- F1=

- Accuracy =

**Scelta iperparametri**

Per la scelta degli iperparametri si è fatto riferimento all’ articolo

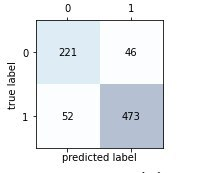
“*Automatic ECG Diagnosis Using Convolutional Neural Network di Roberta Avanzato and Francesco Beritelli* ” e gli iperparametri scelti inizialmente sono stati i seguenti:

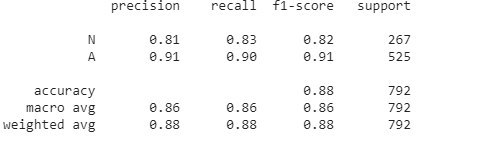


Con questi parametri si sono ottenuti i seguenti risultati

Nella validazione:



****Nel test:

****

Per migliorare i risultati ottenuti sono stati effettuati una serie di test cambiando gli iperparametri della CNN.

I risultati considerati più rilevanti sono riassunti nella seguente tabella.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Testing Result | | | Validation Result | |
| Test | Recall A | Recall N | Accuracy | avg\_accuracy | avg\_recall |
| 1 | 0,99 | 0,40 | 0,80 | 0,834 +/- 0,029 | 0,956 +/- 0,002 |
| 2 | 0,90 | 0,83 | 0,88 | 0,889 +/- 0,023 | 0,968 +/- 0,002 |
| 3 | 0,91 | 0,72 | 0,85 | 0,841 +/- 0,036 | 0,961 +/- 0,002 |
| 4 | 0,94 | 0,77 | 0,89 | 0,861 +/- 0,045 | 0,964 +/- 0,002 |
| 5 | 0,95 | 0,77 | 0,88 | 0,754 +/- 0.096 | 0,901 +/- 0,003 |
| 6 | 0,95 | 0,85 | 0,92 | 0,765 +/- 0,072 | 0,891 +/- 0,003 |
| 7 | 0,98 | 0,51 | 0,83 | 0,845 +/- 0,038 | 0,925 +/- 0,003 |
| 8 | 0,91 | 0,78 | 0,87 | 0,856 +/-0,026 | 0,967 +/- 0,003 |

I test evidenziati sono quelli considerati migliori poiché hanno un recall delle A più alto.

Avere un recall delle A alto significa che la probabilità che un ECG che presenta anomalie venga classificato come normale è molto bassa.

Si è data meno importanza al recall delle N poiché se un ECG normale viene classificato come anomalo è un errore meno grave, poiché si presuppone che gli ECG classificati come anomali vengano successivamente analizzati da un cardiologo.

**Ulteriore Esperimento con segmenti di 15 secondi**

Dividendo i vari tracciati in segmenti di 30 secondi si aveva un eccessivo sbilanciamento del dataset poiché il numero dei segmenti etichettati come anormali era circa il doppio di quelli etichettati come normali ,di conseguenza si è deciso di suddividere i tracciati non più in segmenti di 30 secondi bensì in segmenti di 15 secondi.

In tutto si sono ottenuti i 5280 segmenti di cui:

- 2105 segmenti etichettati come N (40% del dataset)

- 3175 segmenti etichettati come A (60% del dataset)

Anche in questo caso l’intero dataset è stato suddiviso in una parte di learning e una parte di test con una percentuale di 70% per il learning e 30% per il test.

Sia nel learning set che nel test set sono state mantenute le proporzioni iniziali tra segmenti etichettati come A e segmenti etichettati come N.

Eseguendo il test con i gli iperparametri di partenza (seguendo l’articolo precedentemente citato) si sono ottenuti i seguenti risultati:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Testing Result | | | Validation Result | |
| Recall A | Recall N | Accuracy | avg\_accuracy | avg\_recall |
| 0.91 | 0.89 | 0.90 | 0.880 +/- 0.063 | 0.970 +/- 0.002 |

Sono stati effettuati altri test cambiando gli iperparametri e i risultati sono presentati nella seguente tabella.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Testing Result | | | Validation Result | |
| Test | Recall A | Recall N | Accuracy | avg\_accuracy | avg\_recall |
| 1 | 0,88 | 0,86 | 0,88 | 0,784 +/- 0,103 | 0,927 +/- 0,002 |
| 2 | 0,90 | 0,88 | 0,89 | 0,892 +/- 0,022 | 0,971 +/- 0,004 |
| 6 | 1.00 | 0,30 | 0,72 | 0,788 +/- 0,076 | 0,868 +/- 0,002 |
| 7 | 0,99 | 0,38 | 0,75 | 0,728 +/- 0,065 | 0,901 +/- 0,003 |

**Esperimento modificando le proporzioni di A ed N nel TestSet**

Si è provato ad eliminare dal TestSet alcuni segmenti etichettati come A per ottenere una proporzione del 60% di segmenti etichettati come N e il 40% di segmenti etichettati come A, questo per rendere più realistica la fase di Test, poiché precedentemente il test presentava molti più segmenti anomali che normali.

Eseguendo nuovamente il test con gli iperparametri di partenza (seguendo l’ articolo) si sono ottenuti i seguenti risultati:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Testing Result | | | Validation Result | |
| Recall A | Recall N | Accuracy | avg\_accuracy | avg\_recall |
| 0,90 | 0,92 | 0,91 | 0,909 +/- 0,010 | 0,969 +/- 0,003 |

Si sono effettuati successivamente altri test cambiando gli iperparametri e i risultati ottenuti sono presentati nella seguente tabella:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Testing Result | | | Validation Result | |
| Test | Recall A | Recall N | Accuracy | avg\_accuracy | avg\_recall |
| 1 | 0,75 | 0,94 | 0,87 | 0,781 +/- 0,102 | 0,907 +/- 0,002 |
| 4 | 0,91 | 0,85 | 0,88 | 0,887 +/- 0,021 | 0,967+/- 0,001 |
| 6 | 0,80 | 0,96 | 0,89 | 0,821 +/- 0,037 | 0,866 +/- 0,004 |
| 7 | 0,83 | 0,88 | 0,86 | 0,851+/- 0,019 | 0,899 +/- 0,003 |
| 9 | 0,97 | 0,63 | 0,76 | 0,847 +/- 0,045 | 0,926 +/- 0,005 |
| 10 | 1.00 | 0,51 | 0,70 | 0,858 +/- 0,044 | 0,926 +/- 0,003 |
| 12 | 0,90 | 0,91 | 0,91 | 0,864 +/- 0,034 | 0,942 +/- 0,003 |
| 13 | 0,90 | 0,80 | 0,94 | 0,751 +/- 0,083 | 0,900 +/- 0,004 |

I test che hanno ottenuto migliori risultati sono il test 9 e il test 10 poiché presentano un recall delle A più alto.

Analizzando i tracciati classificati male dai due modelli si è considerato come ottimale il test 10, in quanto i soli due errori sono relativi a due tracciati che contengono un'unica extrasistole che tra l'altro si trova una volta in ultima posizione ed una volta in penultima.

Nel test 9, invece oltre ad esserci un maggior numero di tracciati che presentano anomalie, ce ne sono alcuni che ne presentano più di una.

Per la creazione del modello del test numero 10 sono stati scelti i seguenti iperparametri:

-Numero di epoche 50

-Batch size 16

-Pool size 4 per i layer max pooling 1D

-Pool size di 2 per il Layer average pooling 1D

-Kernel size per il primo layer di convoluzione è 35 con strides di 6, per i successivi layer di convoluzione il kernel size è di 4 con striders di 1

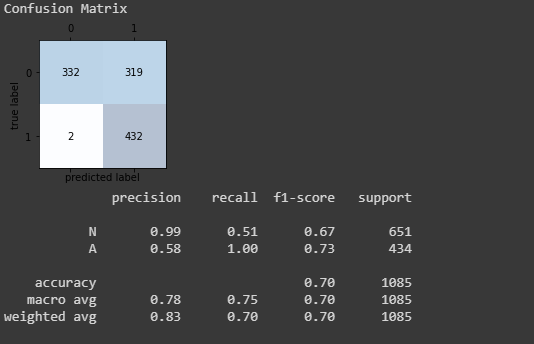
-Funzione di attivazione tanh(tangente iperbolica )

-Funzione di ottimizzazione adamax

-Il layer Dense con funzione di attivazione “softmax” ,kernel inizializer “normal” (inizializzazione dei pesi)

- la dimensione dello spazio di output del primo layer di convoluzione è di 64,i successivi tre layer di convoluzione sono rispettivamente 64,86,120

**Matrice di confusione e precision, recall ed f1-score del test 10**:

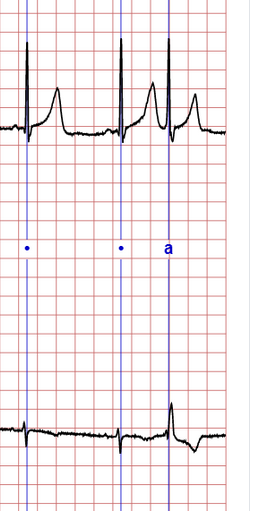


Risultati ottenuti nella validazione con il k-fold:

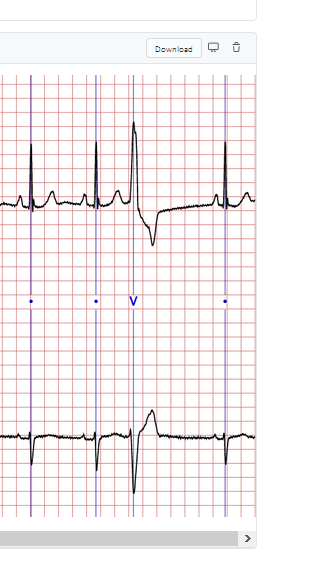
C:\Users\martina\Downloads\val_report.PNG

Tracciati classificati in maniera errata dal modello del test 10:

**Un’unica exatrasistole atriale presente come ultima del tracciato**



**Un’unica exatrasistole ventricolare presente come penultima del tracciato**



**Conclusioni**

I risultati ottenuti sono estremamente soddisfacenti essendo stati classificati come normali soltanto 2 dei 434 ecg che presentavano anomalie  e che in entrambi i casi l'anomalia è presente ad un estrmo della registrazione.  Disponendo di registrazioni  della durata di 30 secondi le probabilità che sia presente una unica anomalia solo al termine della registrazione si dimezzerebbero;  in ogni caso il mancato riconoscimento delle due anomalie presenti dei due casi non ha rilievo dal punto di vista clinico.

L'ottenimento di tale risultato è stato "pagato" con un elevato numero di ecg normali classificati come anomali. Oltre la metà degli ecg normali sono stati classificati correttamente.

Se l'algoritmo potesse essere applicato a tutti gli elettrocardiogrammi che ordinariamente vengono valutati presso i diversi ambulatori di cardiologia, già tale risultato permetterebbe un notevole risparmio.