

Report progetto multimedia signal analysis and understanding

Matteo Mariani 748241

1 Obiettivi del progetto

Il progetto trattato si pone come obiettivo principale l'applicazione di reti neurali per la classificazione automatica dei ritmi cardiaci, regolari ed irregolari, tramite l'utilizzo di segnali ECG, applicando anche il pre-processing illustrato dal paper *ECG Arrhythmia Classification Using STFT-Based Spectrogram and Convolutional Neural Network* [1].

Le aritmie cardiache (i ritmi cardiaci irregolari) rappresentano una delle principali manifestazioni di disfunzione del sistema cardiovascolare. Si tratta di alterazioni del ritmo cardiaco che possono insorgere in forma isolata oppure in concomitanza con altre patologie cardiache. Queste anomalie includono sia battiti accelerati o rallentati, sia contrazioni premature o irregolari, e possono avere conseguenze che spaziano da disturbi lievi fino a condizioni potenzialmente letali.

Per questo progetto sono stati utilizzati i segnali ECG forniti dal database PhysioNet MIT-BIH Arrhythmia. Per affrontare questo problema sono state adottate e sperimentate diverse architetture di reti neurali profonde, tra cui ResNet-18, ResNet-34 e Vision Transformers (ViTs), al fine di valutare l'efficacia e le prestazioni di ciascun modello nella classificazione delle aritmie. Per la classificazione dei ritmi cardiaci sono stati adottati due approcci: il primo consiste nel passare direttamente il segnale ECG alla rete neurale, mentre il secondo approccio consiste nell'applicare lo stesso pre-processing del paper [1] al segnale ECG, che si basa su Short Time Fourier Transforms (STFT) e spettrogrammi.

2 Metodologie per l'analisi dei segnali ECG

L'analisi dei segnali elettrocardiografici (ECG) riveste un ruolo cruciale nella diagnosi automatica delle patologie cardiache, poiché permette di identificare alterazioni del ritmo che possono indicare condizioni cliniche significative. Tradizionalmente, questo processo si articola in due fasi distinte e sequenziali: l'estrazione delle caratteristiche e la successiva classificazione dei pattern.

Nella prima fase, gli studi sul rilevamento si concentrano principalmente sul problema dell'individuazione dei battiti cardiaci nei dati ECG, per cui sono state applicate varie tecniche, tra cui metodi basati su soglie, filtri digitali e trasformate come la Wavelet e di Fourier. La trasformata Wavelet è uno dei metodi più utilizzati per ottenere le caratteristiche dei segnali ECG. Diversi autori hanno proposto tecniche basate su decomposizione Wavelet, algoritmi genetici combinati con reti neurali, estrazione di caratteristiche lineari e non lineari, modelli autoregressivi e altre metodologie.

Dopo l'estrazione delle caratteristiche, la seconda fase del processo tradizionale si concentra sulla classificazione automatica dei pattern identificati. In questo contesto, numerosi algoritmi di machine learning sono stati impiegati per associare le caratteristiche estratte a specifiche classi di aritmie. Le macchine a vettori di supporto (SVM), ad esempio, hanno guadagnato popolarità grazie alla loro capacità di gestire spazi ad alta dimensionalità e di trovare margini di separazione ottimali anche con dataset complessi. Le reti neurali (NN), in particolare le architetture feed-forward con uno o più strati nascosti, sono state utilizzate per apprendere relazioni non lineari tra le caratteristiche e le etichette cliniche. Metodi come l'analisi cluster (CA) hanno permesso di raggruppare battiti simili senza supervisione, mentre algoritmi come le foreste casuali (RF) e l'Optimum-Path Forest (OPF) hanno offerto robustezza contro l'overfitting e capacità di gestire feature eterogenee. In alcuni casi, approcci ibridi hanno combinato tecniche di ottimizzazione, come gli algoritmi genetici, con le reti neurali per affinare la selezione delle caratteristiche e migliorare le prestazioni del classificatore.

Negli ultimi anni, le tecniche di deep learning hanno mostrato prestazioni eccezionali nelle applicazioni di riconoscimento dei pattern. Per questo motivo, i ricercatori hanno rivolto sempre più l'attenzione allo studio della classificazione ECG basata sul deep learning. Sono stati proposti approcci con reti neurali ricorrenti (RNN), reti con memoria a lungo termine (LSTM), e reti neurali convoluzionali (CNN), sia monodimensionali (1D-CNN) che bidimensionali (2D-CNN). Tali studi hanno dimostrato che le reti profonde possono apprendere automaticamente caratteristiche rappresentative complesse direttamente dai dati, riducendo la dipendenza da tecniche manuali di estrazione delle caratteristiche e consentendo lo sviluppo di sistemi end-to-end, capaci di ricevere in input i segnali ECG e fornire in output la classificazione delle aritmie.

Tuttavia, a differenza degli approcci tradizionali, le reti neurali profonde richiedono una quantità considerevole di dati per l'addestramento. Ciò genera un divario tra la dimensione dei dataset disponibili pubblicamente e la necessità di grandi quantità di dati per l'estrazione di caratteristiche profonde.

3 Il Dataset

Il database MIT-BIH Arrhythmia, reso disponibile attraverso la piattaforma PhysioNet, rappresenta uno dei dataset più utilizzati e studiati nella comunità scientifica per lo sviluppo e la validazione di algoritmi di classificazione e rilevamento delle aritmie cardiache. Creato originariamente dal Laboratorio di Ingegneria Biomedica del Massachusetts Institute of Technology in collaborazione con il Beth Israel Hospital di Boston, questo dataset è stato pubblicato nel 1980 con l'obiettivo di fornire una base di riferimento standardizzata per la ricerca nel campo dell'analisi dei segnali elettrocardiografici (ECG). Nel corso degli anni, il MIT-BIH Arrhythmia Database è diventato un punto di riferimento imprescindibile per le ricerche mirate al riconoscimento automatico di anomalie del ritmo cardiaco.

Il MIT-BIH Arrhythmia Database è un sottoinsieme curato di un progetto più ampio chiamato "MIT-BIH Database Project", avviato negli anni '70. Questo progetto originario ha raccolto oltre 4000 ore di registrazioni elettrocardiografiche ambulatoriali (Holter) da pazienti del Beth Israel Hospital. Da questo immenso archivio sono stati selezionati con cura 48 segmenti di 30 minuti ciascuno, scelti per rappresentare la più ampia varietà possibile di condizioni aritmiche, profili demografici diversi e scenari di rumore tipici dell'ambiente clinico. La strategia di selezione mirava a creare un set di dati gestibile ma allo stesso tempo rappresentativo, capace di servire come benchmark affidabile per la validazione di nuovi algoritmi di analisi ECG.

Il database MIT-BIH Arrhythmia contiene registrazioni ECG a due canali provenienti da 48 soggetti, con una durata di 30 minuti ciascuna, annotati beat-by-beat da esperti cardiologi. I dati sono campionati a 360 Hz e comprendono una varietà di ritmi cardiaci: battito normale (normal beat, NOR), battito con blocco di branca sinistra (Left Bundle Branch Block LBB), battito con blocco di branca destra (Right Bundle Branch Block, RBB), contrazione ventricolare prematura (Premature Ventricular Contraction, PVC) e contrazione atriale prematura (Atrial Premature Contraction, APC).

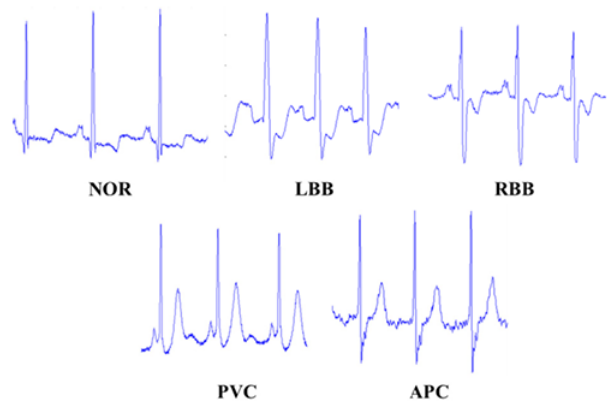


Figure 1: Forme d'onda dei diversi ritmi cardiaci

Ogni record contiene due derivazioni (canali) ECG, con diverse combinazioni di tipologie di derivazioni. In particolare, sono disponibili registrazioni con coppie di derivazioni bipolari e unipolari: MLII-V1, MLII-V2, MLII-V4, MLII-V5 e V5-V2.

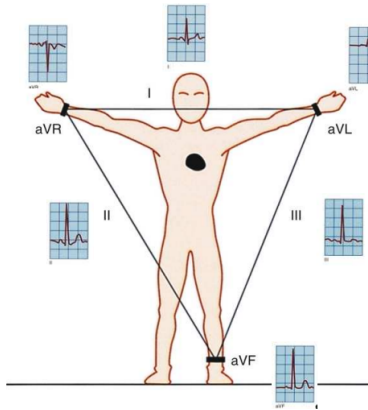


Figure 2: Le derivazioni bipolari

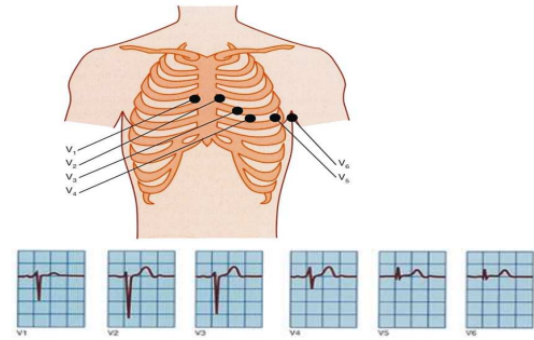


Figure 3: Le derivazioni unipolari

3.1 Acquisizione e selezione dei dati

Per garantire risultati confrontabili, è stata adottata la stessa metodologia dell'articolo *ECG Arrhythmia Classification Using STFT-Based Spectrogram and Convolutional Neural Network* [1]. I campioni per ciascuna categoria di battito sono stati estratti da specifici record del database, utilizzando gli stessi riportati nel paper:

- **NOR**: record 100, 105, 215.
- **LBB**: record 109, 111, 214.
- **RBB**: record 118, 124, 212.
- **APC**: record 207, 209, 232.
- **PVC**: record 106, 233.

Seguendo il paper, ogni registrazione selezionata è stata suddivisa in finestre di durata di 10 secondi. Ottendo complessivamente 2520 finestre ECG a due canali, ovvero finestre composte da 3600 campioni (samples) per canale. A partire da tali finestre a due canali, sono stati poi estratti i singoli canali, ottenendo così $2520 \cdot 2 = 5040$ finestre ECG ad un canale.

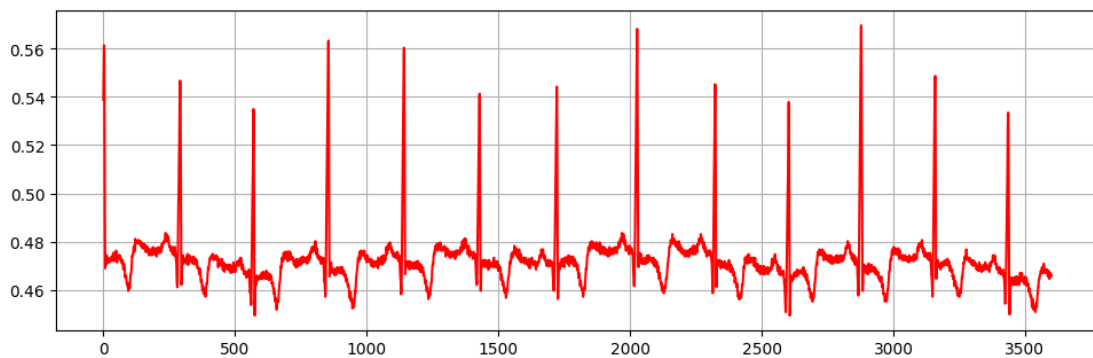


Figure 4: Esempio di una finestra

Il paper fornisce anche il dettaglio numerico della suddivisione training/testing per categoria: per le quattro categorie NOR, LBB, RBB e APC l'articolo indica 900 campioni per il training e 180 campioni per il testing (per ciascuna di queste classi), mentre per la classe PVC sono stati usati 600 campioni per il training e 120 per il testing. In altre parole, la ripartizione adottata equivale ad un rapporto approssimativo di 0.83 per il training e di 0.16 per il testing.

3.2 Pre-processing

Per la fase di pre-processing è stata seguita la stessa procedura dall'articolo [1], con l'aggiunta della specifica relativa alla scelta dell'hop size. Ogni finestra di 10s del segnale ECG, è stata trasformata dal dominio del tempo a quello tempo-frequenza mediante la *Short-Time Fourier Transform* (STFT), definita come:

$$STFT\{x[n]\} = X(m, \omega) = \sum_{n=-\infty}^{\infty} x[n] w[n-m] e^{-j\omega n}$$

dove:

- $w[n]$ è la finestra di analisi,
- m è lo shift della finestra,
- ω è la pulsazione angolare in rad/s.

Per la finestra $w[n]$ è stata utilizzata la finestra di Hanning:

$$w[n] = \begin{cases} 0.5 \left[1 - \cos \left(\frac{2\pi n}{M-1} \right) \right], & 0 \leq n \leq M-1 \\ 0, & \text{otherwise} \end{cases}$$

Questa finestra permette di ridurre il leakage spettrale. Le transizioni ai bordi della finestra sono dolci (iniziano e finiscono a zero). Questo riduce le discontinuità tra diversi frame.

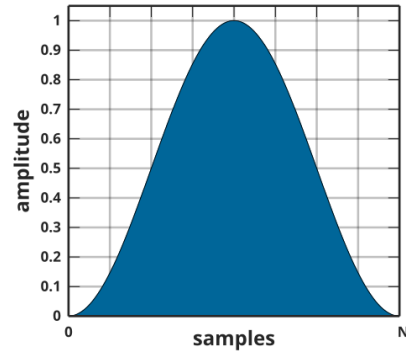


Figure 5

Nel paper [1], la dimensione della finestra utilizzata è pari a $M = 512$ campioni. Poiché non viene specificato l'hop size, in questa implementazione è stato scelto un valore pari al 50% della finestra. In questo modo, le finestre risultano sovrapposte del 50%, garantendo una buona risoluzione temporale senza perdere informazione frequenziale.

Dalla STFT, lo spettrogramma è stato calcolato come modulo al quadrato dei coefficienti:

$$S(m, \omega) = |X(m, \omega)|^2 \quad (1)$$

Infine, ogni spettrogramma è stato trasformato in scala logaritmica.

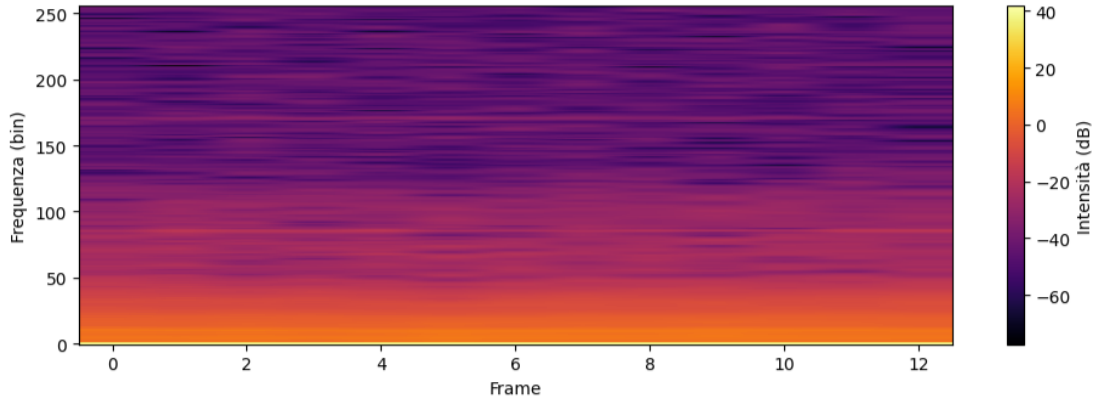


Figure 6: Esempio di spettrogramma in scala logaritmica

Successivamente, lo spettrogramma è stato normalizzato e ridimensionato ad un'immagine di dimensione 256×256 pixel, che costituisce l'input per le diverse reti neurali.

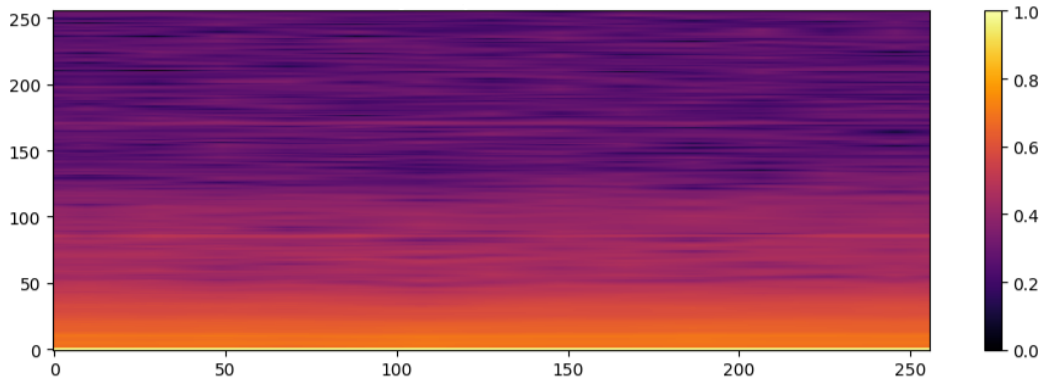


Figure 7: Esempio di spettrogramma ridimensionato e normalizzato

4 Addestramento delle reti neurali

4.1 Modelli utilizzati

In questo progetto sono state impiegate diverse architetture di reti neurali: ResNet18, ResNet34, Vision Transformer (ViT) e la convolutional neural network (CNN) proposta in [1]. L'obiettivo è stato quello di confrontare le prestazioni di ciascun modello sia con che senza l'applicazione del pre-processing sui segnali ECG, descritto in precedenza.

Nei casi in cui non è stato utilizzato il pre-processing, le architetture sono state opportunamente adattate per accettare immagini 1D (vettori), dal momento che tali modelli sono originariamente progettati per operare su immagini 2D (matrici).

Per l'architettura ViT sono stati scelti i seguenti parametri: `kernel_size = 16`, `kernel_stride = 16`, `kernel_in_channels = 1`, `embed_dim = 256`, `num_layers = 6`, `num_heads = 4`, `mlp_dim = 256 × 3`, `dropout = 0.1`.

Nella versione del ViT applicata a immagini 1D, i parametri `kernel_size` e `kernel_stride` sono stati modificati a 72, in quanto l'architettura opera su un vettore di 3600 elementi.

4.2 Elementi e parametri adottati per il training

L'addestramento di ciascun modello è stato ripetuto per tre volte, utilizzando ad ogni iterazione un *seed* diverso, al fine di suddividere casualmente i dati del dataset in training e validation. In ciascuna iterazione, il training è stato eseguito per un totale di 100 epoche, con un *batch size* pari a 150 elementi.

Per il controllo del tasso di apprendimento è stato impiegato lo scheduler *CosineAnnealingLR*, con un valore iniziale di *learning rate* pari a $5 \cdot 10^{-4}$ e un valore minimo $\eta_{\min} = 2.5 \cdot 10^{-5}$.

L'ottimizzazione dei pesi è stata effettuata utilizzando l'algoritmo *Adagrad*, con un valore di weight decay pari a $1 \cdot 10^{-3}$.

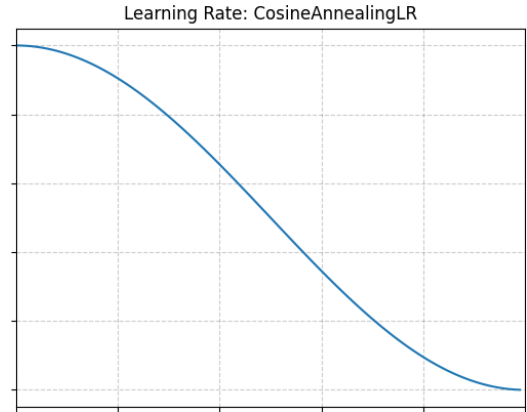


Figure 8: Grafico CosineAnnealingLR

Come funzione di perdita è stata adottata la *focal cross-entropy*, definita come:

$$\text{FL}(y, \hat{p}) = - \sum_{i=1}^n y_i \alpha_i (1 - \hat{p}_i)^\gamma \log(\hat{p}_i),$$

dove γ rappresenta il parametro di *focusing*, fissato a 2 in questo caso. La *cross-entropy*, definita come

$$\text{CE}(y, \hat{p}) = - \sum_{i=1}^n y_i \log(\hat{p}_i).$$

si differenzia dalla *focal cross-entropy* per l'aggiunta del termine $(1 - \hat{p}_i)^\gamma$, che riduce il contributo degli esempi classificati correttamente, concentrando l'apprendimento su quelli più difficili.

4.3 Risultati del training

Di seguito sono riportati i risultati ottenuti durante il processo di training di ciascun modello.

Modelli con STFT

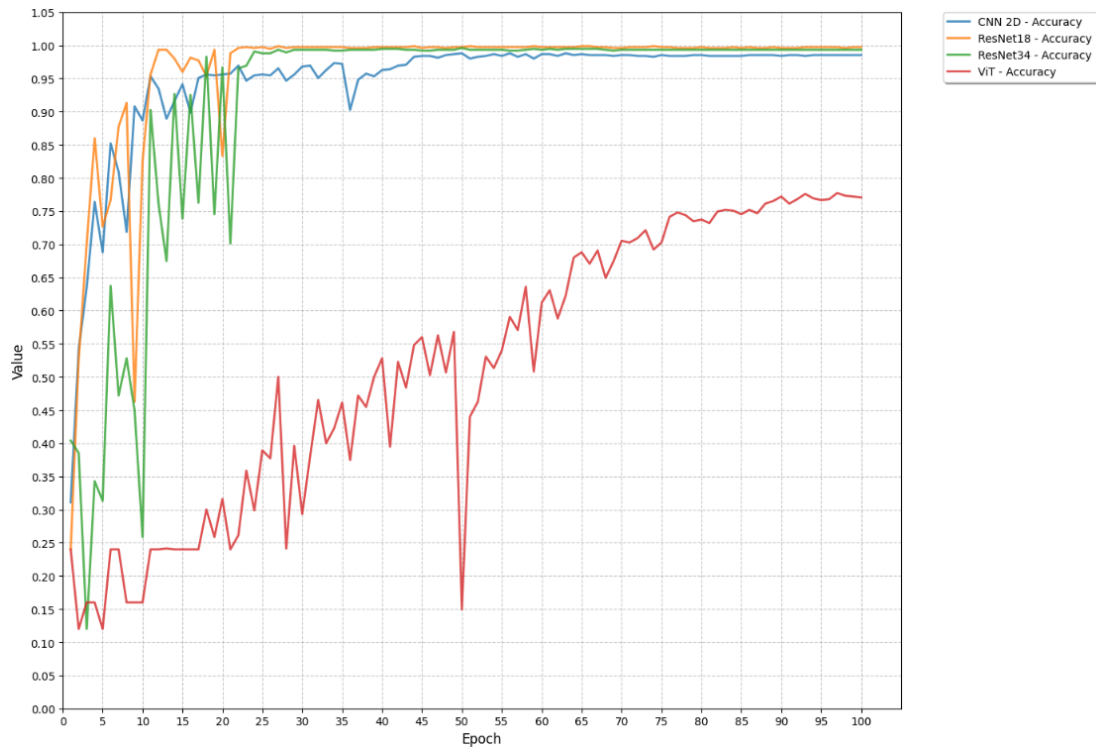


Figure 9: seed 48

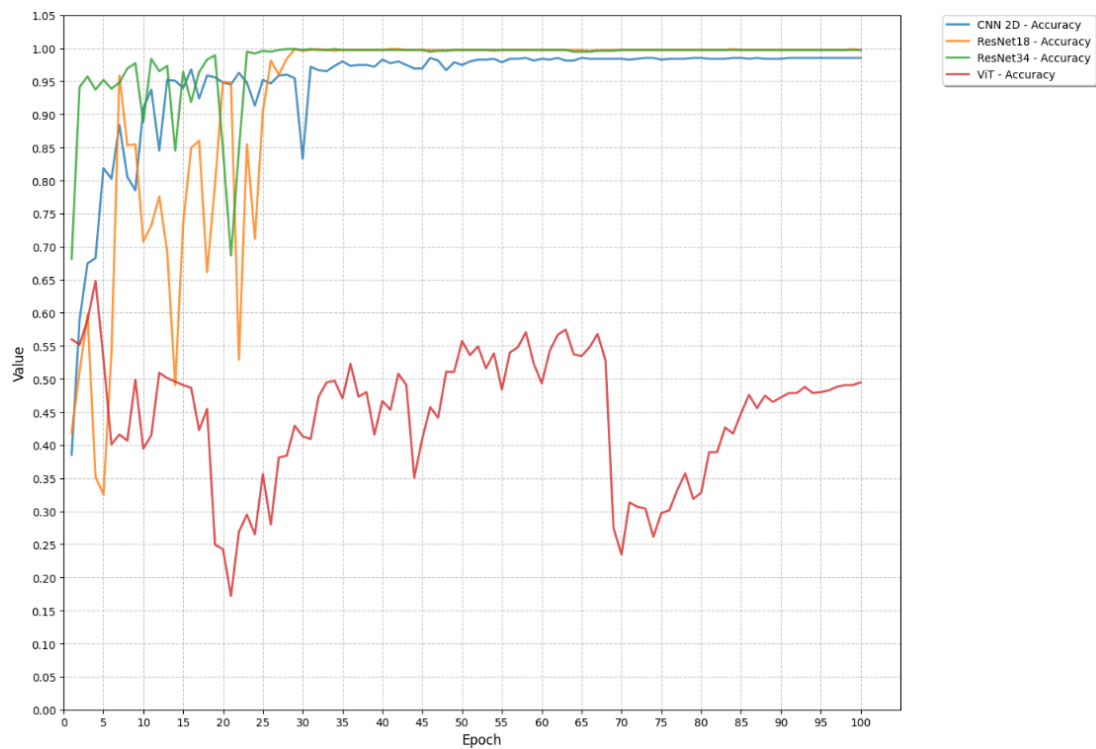


Figure 10: seed 2304

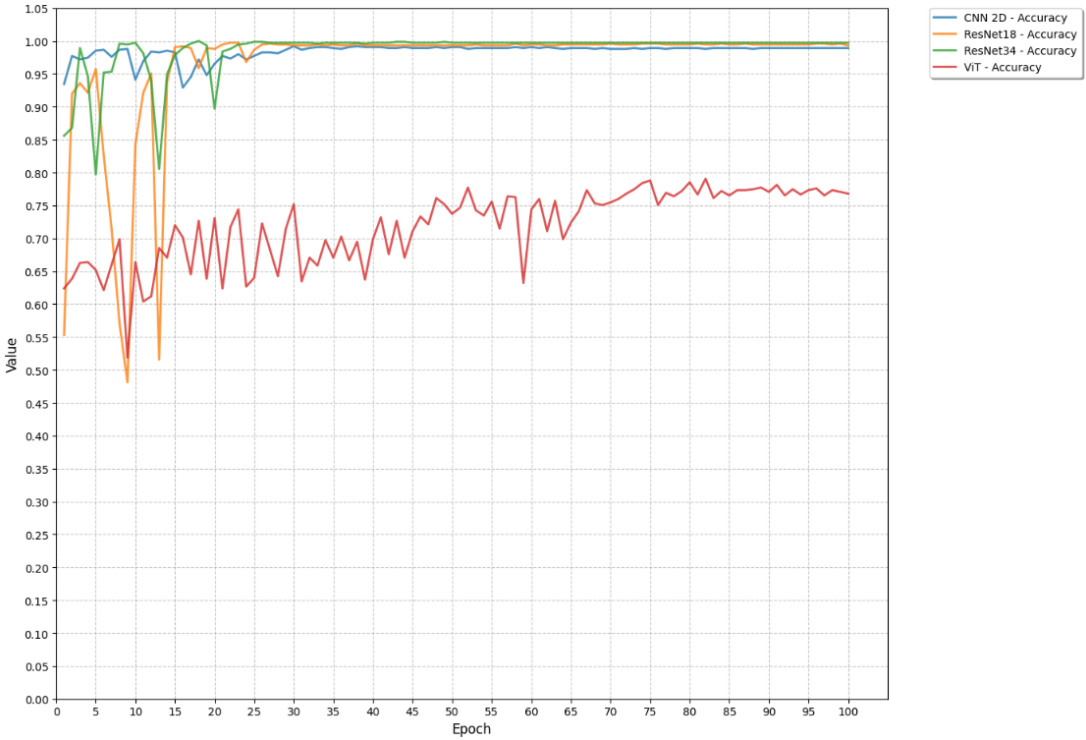


Figure 11: seed 110592

Modelli senza STFT

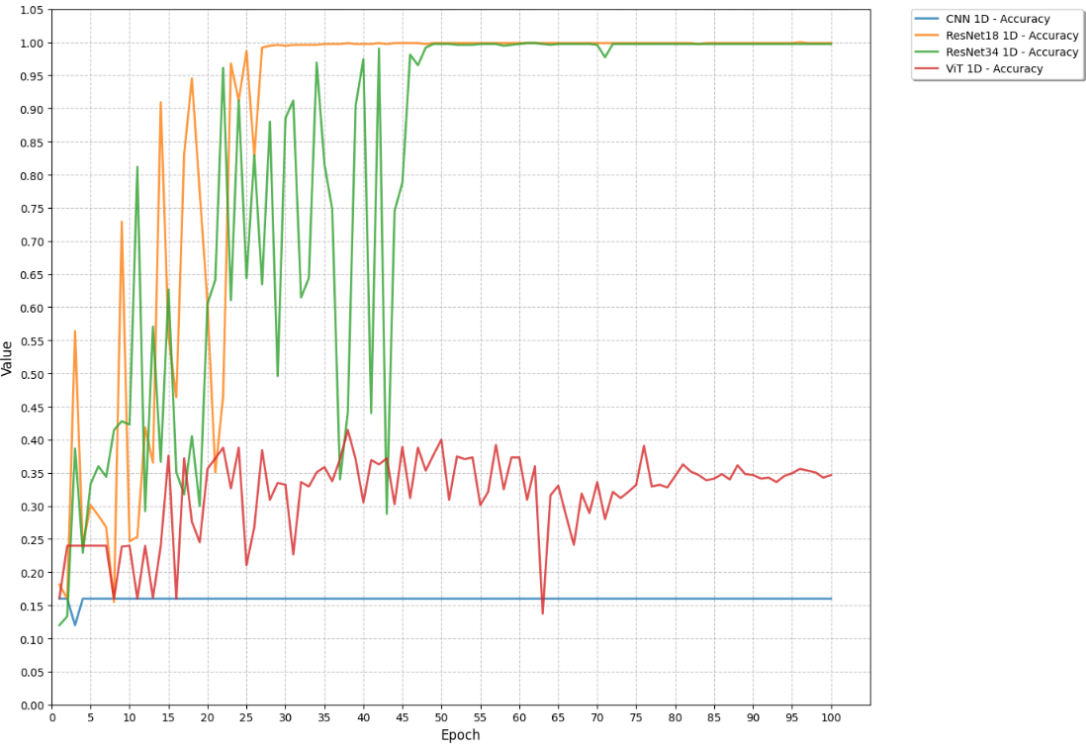


Figure 12: seed 48

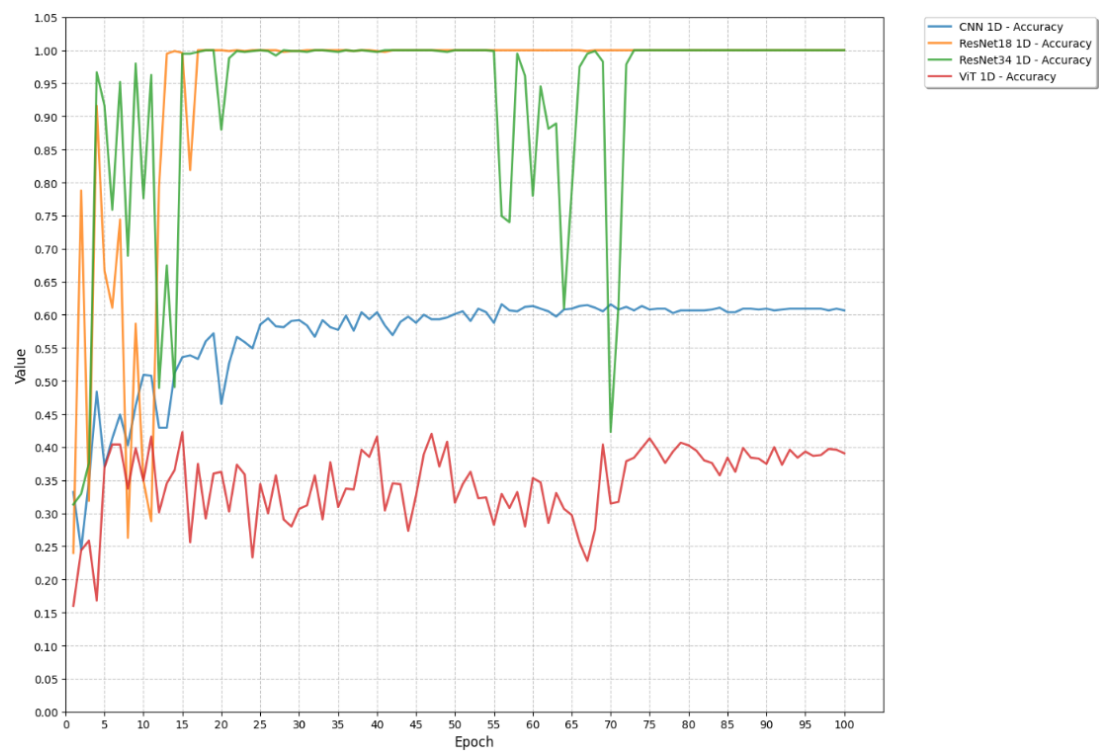


Figure 13: seed 2304

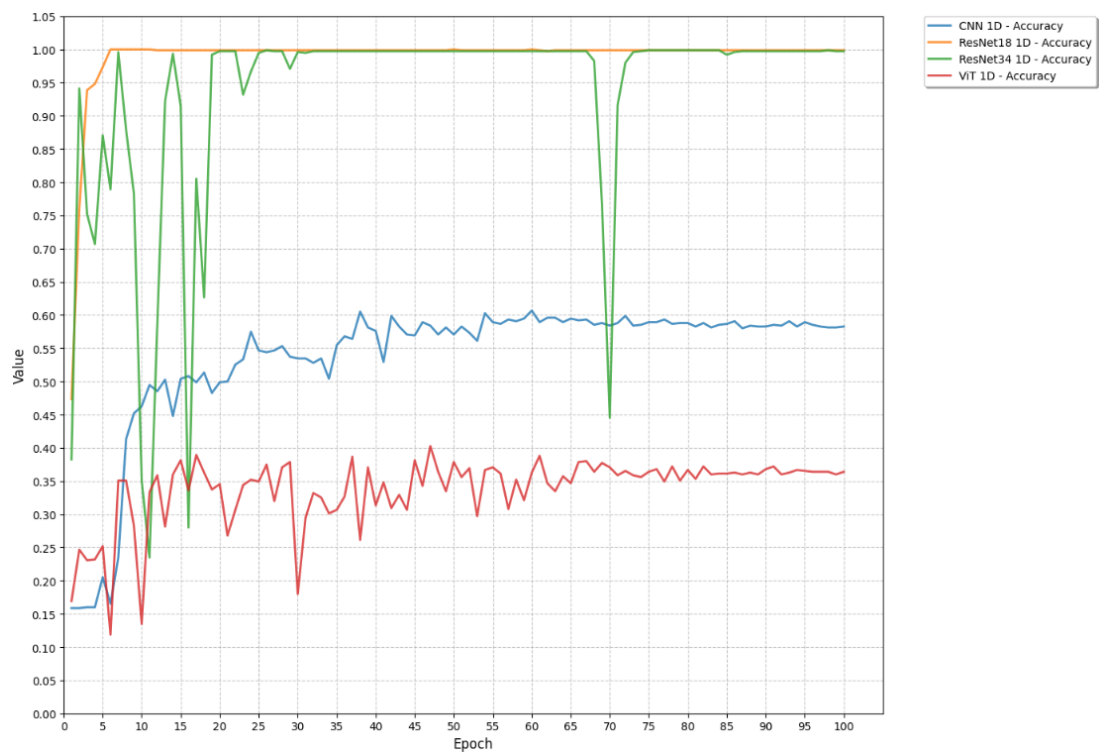


Figure 14: seed 110592

4.4 Matrici di confusione

CNN 1D

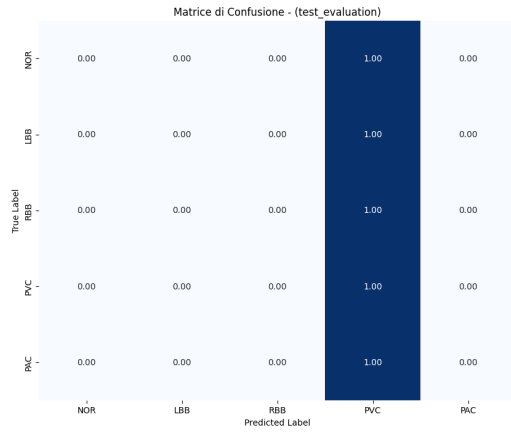


Figure 15

CNN 2D + STFT

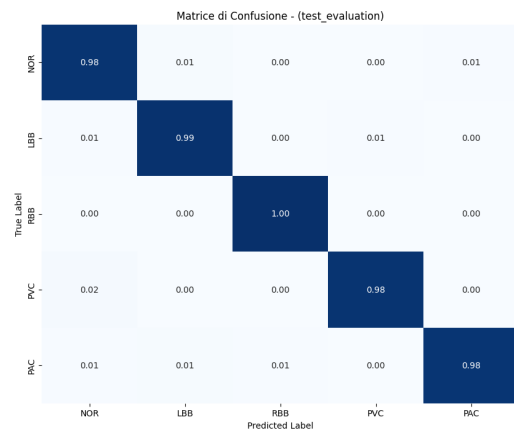


Figure 18

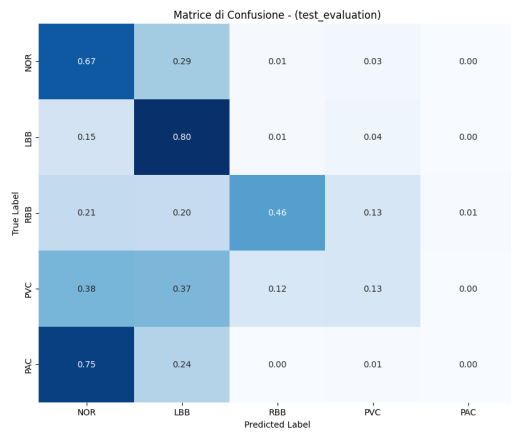


Figure 16

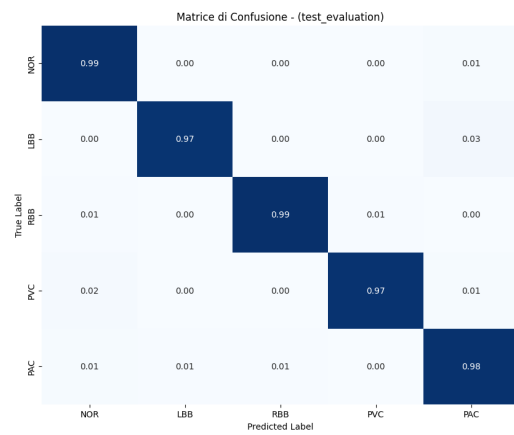


Figure 19

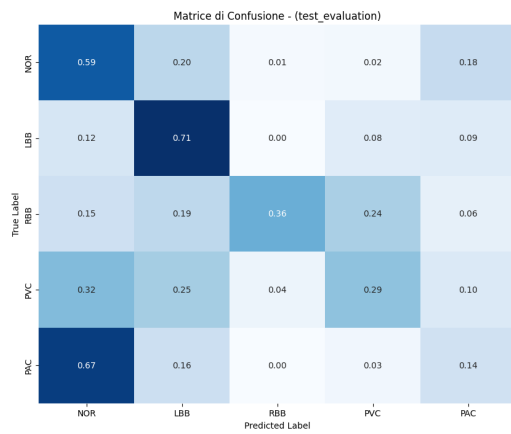


Figure 17

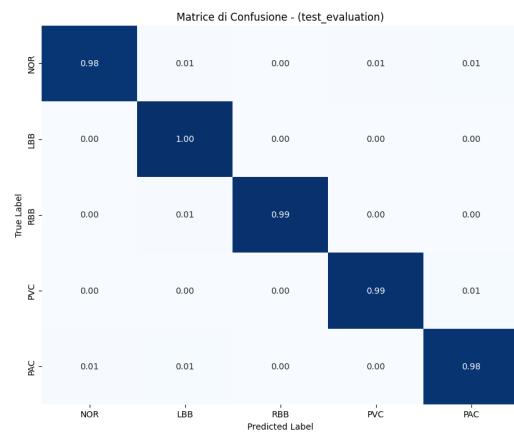


Figure 20

ResNet18 1D

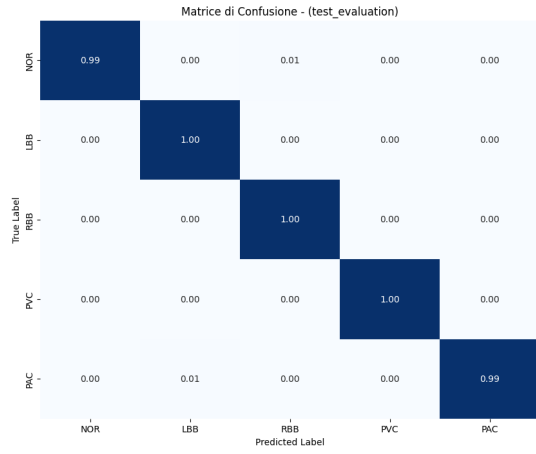


Figure 21

ResNet18 2D + STFT

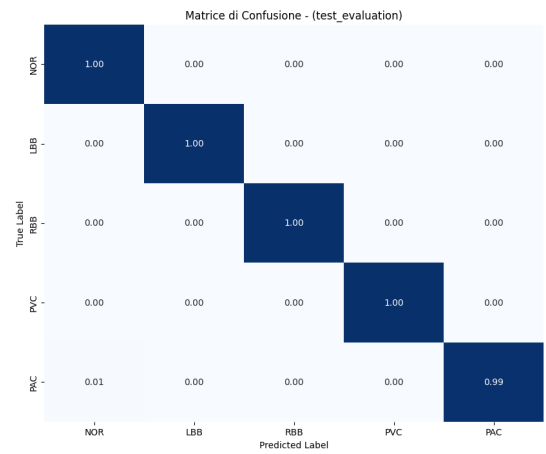


Figure 24

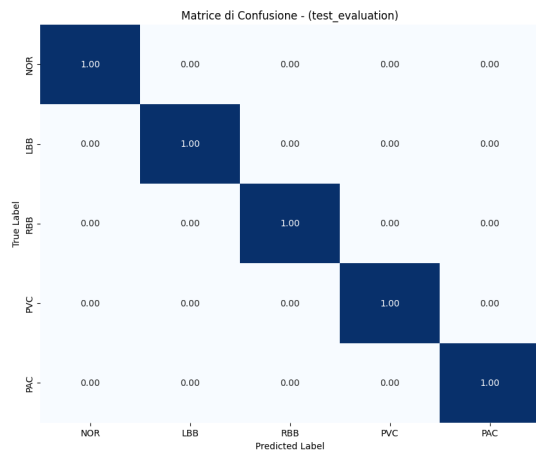


Figure 22

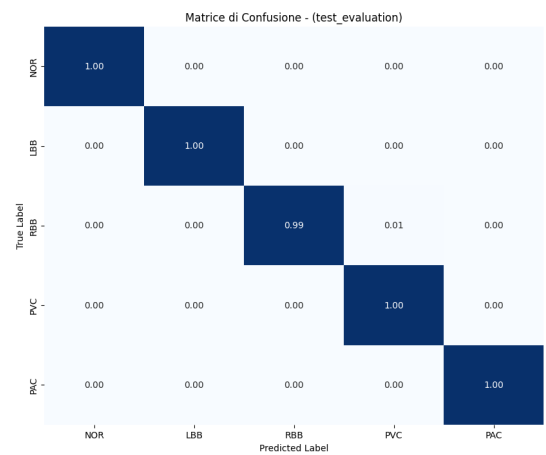


Figure 25

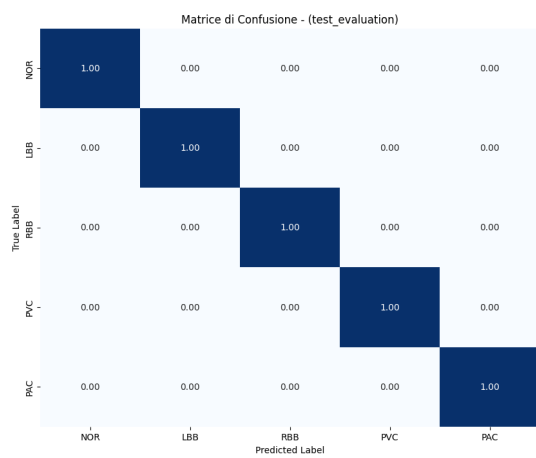


Figure 23

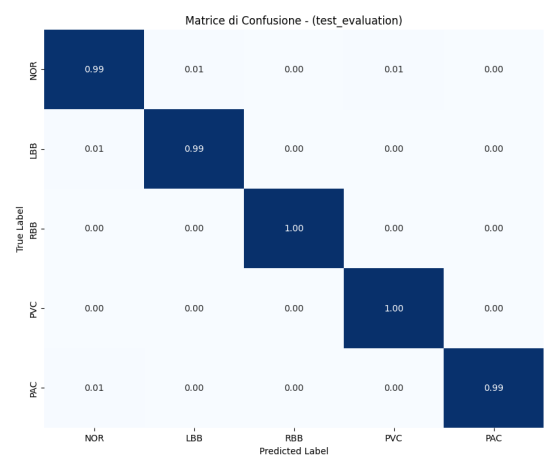


Figure 26

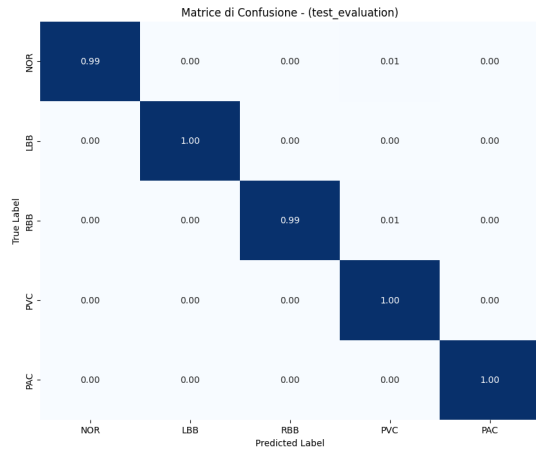
ResNet34 1D

Figure 27

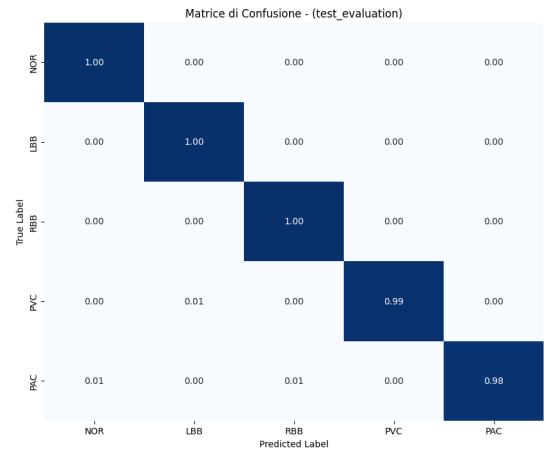
ResNet34 2D + STFT

Figure 30

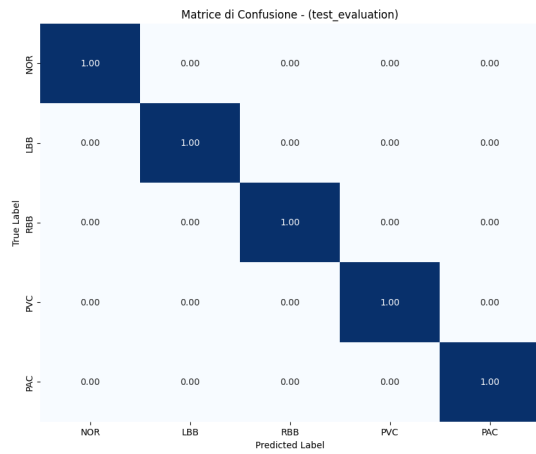


Figure 28

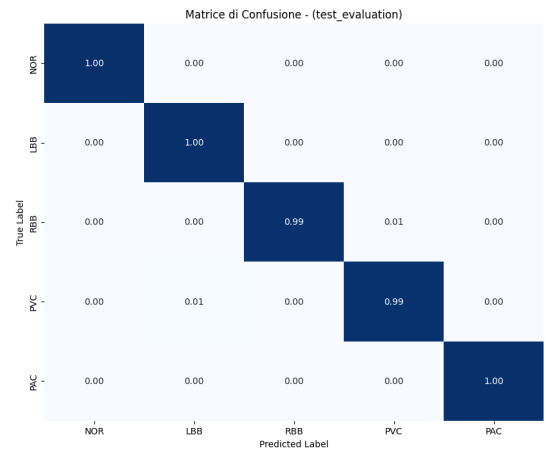


Figure 31

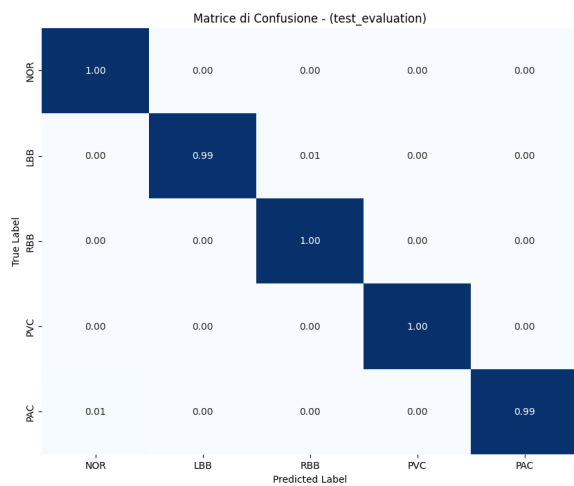


Figure 29

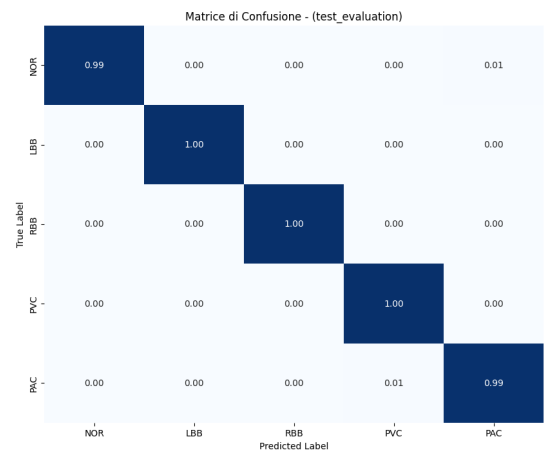


Figure 32

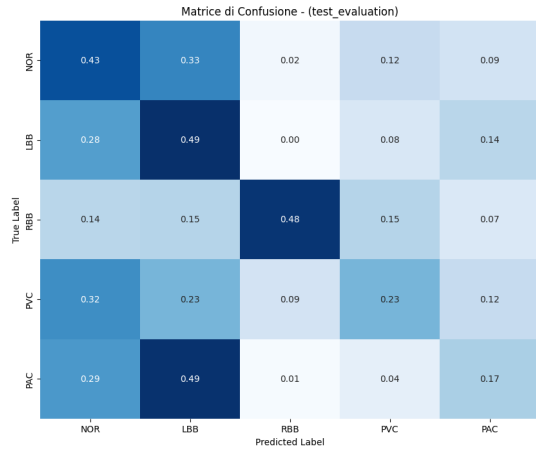
ViT 1D

Figure 33

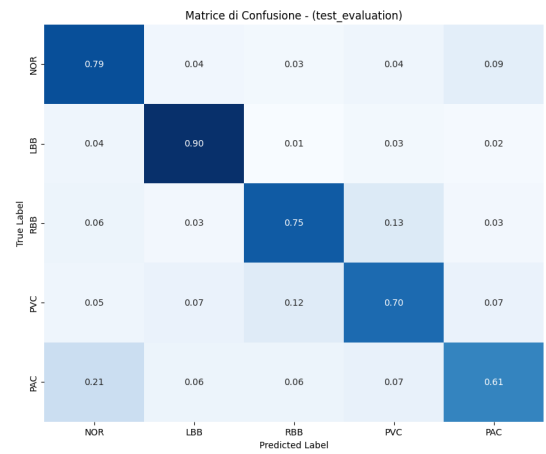
ViT 2D + STFT

Figure 36

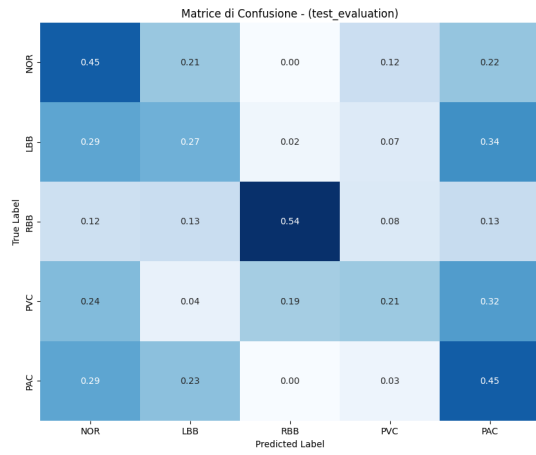


Figure 34

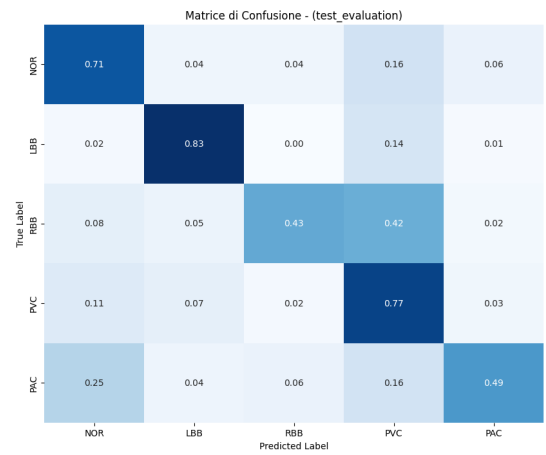


Figure 37

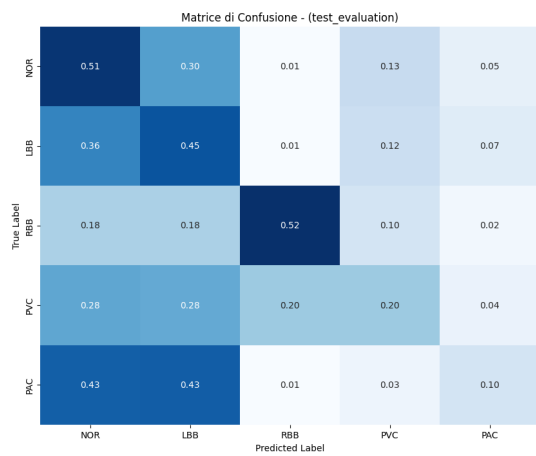


Figure 35

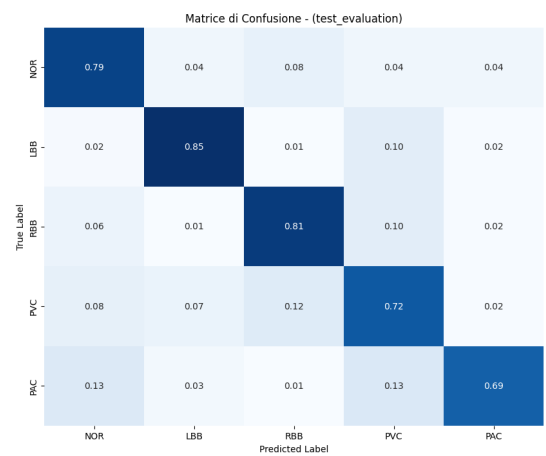


Figure 38

5 Risultati ottenuti

Per la valutazione di ciascun modello è stata utilizzata la versione corrispondente all'epoca in cui è stato registrato il valore minimo di *loss*.

5.1 Modelli 2D + STFT

Categoria	seed 48				seed 2304				seed 110592			
	Istanze	Precision	Recall	F1	Istanze	Precision	Recall	F1	Istanze	Precision	Recall	F1
NOR	180	0.9779	0.9833	0.9806	180	0.9727	0.9889	0.9807	180	0.9944	0.9778	0.9860
LBB	180	0.9780	0.9889	0.9834	180	0.9943	0.9722	0.9831	180	0.9783	1.0000	0.9890
RBB	180	0.9945	1.0000	0.9972	180	0.9944	0.9889	0.9916	180	1.0000	0.9944	0.9972
PVC	120	0.9916	0.9833	0.9874	120	0.9915	0.9750	0.9832	120	0.9917	0.9917	0.9917
PAC	180	0.9944	0.9778	0.9860	180	0.9565	0.9778	0.9670	180	0.9833	0.9833	0.9833
Weighted Avg		0.9870	0.9869	0.9869		0.9812	0.9810	0.9810		0.9894	0.9893	0.9893
Overall Accuracy	0.9869				0.9810				0.9893			
Overall Kappa	0.9835				0.9761				0.9865			

Table 1: Risultati di Valutazione per CNN 2D del paper

Categoria	seed 48				seed 2304				seed 110592			
	Istanze	Precision	Recall	F1	Istanze	Precision	Recall	F1	Istanze	Precision	Recall	F1
NOR	180	0.9945	1.0000	0.9972	180	1.0000	1.0000	1.0000	180	0.9889	0.9889	0.9889
LBB	180	1.0000	1.0000	1.0000	180	1.0000	1.0000	1.0000	180	0.9944	0.9944	0.9944
RBB	180	1.0000	1.0000	1.0000	180	1.0000	0.9944	0.9972	180	1.0000	1.0000	1.0000
PVC	120	1.0000	1.0000	1.0000	120	0.9917	1.0000	0.9959	120	0.9917	1.0000	0.9959
PAC	180	1.0000	0.9944	0.9972	180	1.0000	1.0000	1.0000	180	1.0000	0.9944	0.9972
Weighted Avg		0.9988	0.9988	0.9988		0.9988	0.9988	0.9988		0.9952	0.9952	0.9952
Overall Accuracy	0.9988				0.9988				0.9952			
Overall Kappa	0.9985				0.9985				0.9940			

Table 2: Risultati di Valutazione per ResNet18

Categoria	seed 48				seed 2304				seed 110592			
	Istanze	Precision	Recall	F1	Istanze	Precision	Recall	F1	Istanze	Precision	Recall	F1
NOR	180	0.9890	1.0000	0.9945	180	1.0000	1.0000	1.0000	180	1.0000	0.9944	0.9972
LBB	180	0.9945	1.0000	0.9972	180	0.9945	1.0000	0.9972	180	1.0000	1.0000	1.0000
RBB	180	0.9945	1.0000	0.9972	180	1.0000	0.9944	0.9972	180	1.0000	1.0000	1.0000
PVC	120	1.0000	0.9917	0.9958	120	0.9917	0.9917	0.9917	120	0.9917	1.0000	0.9959
PAC	180	1.0000	0.9833	0.9916	180	1.0000	1.0000	1.0000	180	0.9944	0.9944	0.9944
Weighted Avg		0.9953	0.9952	0.9952		0.9976	0.9976	0.9976		0.9976	0.9976	0.9976
Overall Accuracy	0.9952				0.9976				0.9976			
Overall Kappa	0.9940				0.9970				0.9970			

Table 3: Risultati di Valutazione per ResNet34

Categoria	seed 48				seed 2304				seed 110592			
	Istanze	Precision	Recall	F1	Istanze	Precision	Recall	F1	Istanze	Precision	Recall	F1
NOR	180	0.7010	0.7944	0.7448	180	0.6287	0.7056	0.6649	180	0.7474	0.7889	0.7676
LBB	180	0.8394	0.9000	0.8686	180	0.8187	0.8278	0.8232	180	0.8644	0.8500	0.8571
RBB	180	0.8182	0.7500	0.7826	180	0.8041	0.4333	0.5632	180	0.8202	0.8111	0.8156
PVC	120	0.6316	0.7000	0.6640	120	0.3680	0.7667	0.4973	120	0.5584	0.7167	0.6277
PAC	180	0.7586	0.6111	0.6769	180	0.8165	0.4944	0.6159	180	0.8865	0.6944	0.7788
Weighted Avg		0.7582	0.7548	0.7534		0.7100	0.6369	0.6426		0.7909	0.7762	0.7795
Overall Accuracy	0.7548				0.6369				0.7762			
Overall Kappa	0.6923				0.5501				0.7198			

Table 4: Risultati di Valutazione per ViT

5.2 Modelli 1D

Categoria	SEED 48				SEED 2304				SEED 110592			
	Istanze	Precision	Recall	F1	Istanze	Precision	Recall	F1	Istanze	Precision	Recall	F1
NOR	180	0.0000	0.0000	0.0000	180	0.3315	0.6722	0.4440	180	0.3408	0.5944	0.4332
LBB	180	0.0000	0.0000	0.0000	180	0.4500	0.8000	0.5760	180	0.4981	0.7111	0.5858
RBB	180	0.0000	0.0000	0.0000	180	0.8119	0.4556	0.5836	180	0.9014	0.3556	0.5100
PVC	120	0.1429	1.0000	0.2500	120	0.3019	0.1333	0.1850	120	0.3465	0.2917	0.3167
PAC	180	0.0000	0.0000	0.0000	180	0.0000	0.0000	0.0000	180	0.2680	0.1444	0.1877
Weighted Avg		0.0204	0.1429	0.0357		0.3846	0.4321	0.3701		0.4798	0.4286	0.4131
Overall Accuracy	0.1429				0.4321				0.4286			
Overall Kappa	0.0000				0.2814				0.2806			

Table 5: Risultati di Valutazione per CNN 1D

Categoria	SEED 48				SEED 2304				SEED 110592			
	Istanze	Precision	Recall	F1	Istanze	Precision	Recall	F1	Istanze	Precision	Recall	F1
NOR	180	1.0000	0.9944	0.9972	180	1.0000	1.0000	1.0000	180	0.9889	0.9889	0.9889
LBB	180	0.9945	1.0000	0.9972	180	1.0000	1.0000	1.0000	180	0.9944	0.9944	0.9944
RBB	180	0.9945	1.0000	0.9972	180	1.0000	1.0000	1.0000	180	1.0000	1.0000	1.0000
PVC	120	1.0000	1.0000	1.0000	120	1.0000	1.0000	1.0000	120	0.9917	1.0000	0.9959
PAC	180	1.0000	0.9944	0.9972	180	1.0000	1.0000	1.0000	180	1.0000	0.9944	0.9972
Weighted Avg		0.9976	0.9976	0.9976		1.0000	1.0000	1.0000		0.9952	0.9952	0.9952
Overall Accuracy	0.9976				1.0000				0.9952			
Overall Kappa	0.9970				1.0000				0.9940			

Table 6: Risultati di Valutazione per ResNet18 1D

Categoria	SEED 48				SEED 2304				SEED 110592			
	Istanze	Precision	Recall	F1	Istanze	Precision	Recall	F1	Istanze	Precision	Recall	F1
NOR	180	1.0000	0.9944	0.9972	180	1.0000	1.0000	1.0000	180	0.9890	1.0000	0.9945
LBB	180	1.0000	1.0000	1.0000	180	1.0000	1.0000	1.0000	180	1.0000	0.9944	0.9972
RBB	180	1.0000	0.9944	0.9972	180	1.0000	1.0000	1.0000	180	0.9945	1.0000	0.9972
PVC	120	0.9836	1.0000	0.9917	120	1.0000	1.0000	1.0000	120	1.0000	1.0000	1.0000
PAC	180	1.0000	1.0000	1.0000	180	1.0000	1.0000	1.0000	180	1.0000	0.9889	0.9944
Weighted Avg		0.9977	0.9976	0.9976		1.0000	1.0000	1.0000		0.9965	0.9964	0.9964
Overall Accuracy	0.9976				1.0000				0.9964			
Overall Kappa	0.9970				1.0000				0.9955			

Table 7: Risultati di Valutazione per ResNet34 1D

Categoria	SEED 48				SEED 2304				SEED 110592			
	Istanze	Precision	Recall	F1	Istanze	Precision	Recall	F1	Istanze	Precision	Recall	F1
NOR	180	0.3171	0.4333	0.3662	180	0.3418	0.4500	0.3885	180	0.3077	0.5111	0.3841
LBB	180	0.3048	0.4944	0.3771	180	0.3097	0.2667	0.2866	180	0.2914	0.4500	0.3537
RBB	180	0.8529	0.4833	0.6170	180	0.7840	0.5444	0.6426	180	0.7769	0.5222	0.6246
PVC	120	0.2828	0.2333	0.2557	120	0.3165	0.2083	0.2513	120	0.2553	0.2000	0.2243
PAC	180	0.2970	0.1667	0.2135	180	0.3320	0.4500	0.3821	180	0.3750	0.1000	0.1579
Weighted Avg		0.4201	0.3714	0.3738		0.4239	0.3964	0.4001		0.4117	0.3679	0.3578
Overall Accuracy	0.3714				0.3964				0.3679			
Overall Kappa	0.2085				0.2383				0.2036			

Table 8: Risultati di Valutazione per Vit 1D

5.3 Riepilogo e confronto dei risultati

Modello	Seed 48		Seed 2304		Seed 110592		Media	
	Acc	Kappa	Acc	Kappa	Acc	Kappa	Acc	Kappa
CNN 2D + STFT	0.9869	0.9835	0.9810	0.9761	0.9893	0.9865	0.9857	0.9820
ResNet18 + STFT	0.9988	0.9985	0.9988	0.9985	0.9952	0.9940	0.9976	0.9970
ResNet34 + STFT	0.9952	0.9940	0.9976	0.9970	0.9976	0.9970	0.9968	0.9960
ViT + STFT	0.7548	0.6923	0.6369	0.5501	0.7762	0.7198	0.7226	0.6541
CNN 1D	0.1429	0.0000	0.4321	0.2814	0.4286	0.2806	0.3345	0.1873
ResNet18 1D	0.9976	0.9970	1.0000	1.0000	0.9952	0.9940	0.9976	0.9970
ResNet34 1D	0.9976	0.9970	1.0000	1.0000	0.9964	0.9955	0.9980	0.9975
ViT 1D	0.3714	0.2085	0.3964	0.2383	0.3679	0.2036	0.3786	0.2168
CNN 2D paper	-	-	-	-	-	-	0.99.00	-
CNN 1D paper	-	-	-	-	-	-	0.90.00	-

Table 9: Riepilogo Overall Accuracy e Overall Kappa per modello

6 Analisi dei Risultati

L'analisi dei risultati ottenuti evidenzia differenze significative tra le diverse architetture e rappresentazioni dei segnali ECG.

I modelli **basati su trasformata STFT e convoluzioni 2D** hanno confermato quanto riportato nel paper [1], mostrando ottime prestazioni in termini di accuratezza e coefficiente kappa. In particolare, la **ResNet18** e la **ResNet34** hanno raggiunto valori prossimi alla perfezione (Overall Accuracy ≥ 0.995 e Kappa ≥ 0.994), superando in alcuni casi i risultati della CNN 2D standard implementata dagli autori. Questo conferma come l'utilizzo di architetture più profonde e pre-addestrate riesca a catturare meglio le caratteristiche tempo-frequenziali dello spettrogramma, con una generalizzazione superiore rispetto a reti più semplici.

Il **ViT con input STFT** ha invece evidenziato performance molto inferiori (Accuracy media ≈ 0.72 , Kappa ≈ 0.65). Nonostante il Transformer sia in grado di modellare relazioni di lungo raggio, i risultati suggeriscono che per segnali ECG la rappresentazione spettrale bidimensionale è meglio sfruttata da reti convoluzionali, probabilmente per via della struttura locale e ripetitiva delle forme d'onda. Inoltre, il dataset relativamente limitato potrebbe non essere sufficiente per addestrare in modo efficace un modello Transformer.

Passando ai modelli **1D**, emerge una netta distinzione:

- La **CNN 1D semplice** non ha raggiunto prestazioni accettabili (Accuracy media ≈ 0.33 , Kappa ≈ 0.18), risultando sensibilmente peggiore rispetto alla CNN 1D del paper di riferimento (≈ 0.90). Ciò potrebbe dipendere da una configurazione non ottimale dell'architettura o da un'estrazione insufficiente di pattern discriminativi dalle sole sequenze temporali.
- Le varianti **ResNet18 1D** e **ResNet34 1D** hanno mostrato risultati eccellenti (Accuracy media ≥ 0.997 , Kappa ≥ 0.997), addirittura superiori ai modelli 2D+STFT. Questo indica che architetture residuali profonde sono in grado di apprendere direttamente dal segnale grezzo informazioni discriminative di elevata qualità, senza dover passare da una trasformata tempo-frequenza.
- Il **ViT 1D**, invece, ha performato male (Accuracy media ≈ 0.37 , Kappa ≈ 0.22), analogamente al ViT 2D. Questo conferma la difficoltà dei Transformer ad apprendere in assenza di dataset molto ampi o di strategie di pre-addestramento specifiche.

Nel complesso, i risultati mostrano che:

1. le **ResNet** (sia 1D che 2D+STFT) sono le architetture migliori per la classificazione delle aritmie ECG;
2. le **CNN tradizionali** funzionano bene solo con rappresentazioni spettrali, ma non sono sufficientemente potenti in versione 1D;
3. i **Transformer (ViT)**, pur promettenti, non risultano adatti in questo scenario senza opportune strategie di ottimizzazione e maggiori quantità di dati.

Infine, il confronto con i risultati del paper [1] evidenzia che i modelli proposti hanno non solo replicato le prestazioni della CNN 2D originaria, ma in alcuni casi le hanno anche superate, soprattutto grazie all'introduzione delle architetture residuali.

7 Articoli

- [1] J. Huang, B. Chen, B. Yao and W. He, "ECG Arrhythmia Classification Using STFT-Based Spectrogram and Convolutional Neural Network," in IEEE Access, vol. 7, pp. 92871-92880, 2019, doi: 10.1109/ACCESS.2019.2928017.