

#### 4.5.3 Risultati rappresentazione con etichette semantiche

Nel corrente paragrafo andremo ad analizzare la seconda fase di classificazione che analizza i risultati dei problemi rappresentati da categorie semantiche. Come precedentemente affermato, la situazione in cui si sono disegnate queste casistiche è un condizione multi-etichetta. È possibile presupporre che ogni risultato sia stato influenzato negativamente dalla presenza di altri suoni non considerati nel problema in esame, come, allo stesso tempo, la quantità limitata di dati a disposizione per classe potrebbe conseguire un scarso addestramento condizionando la capacità di generalizzazione del discriminatore. Tuttavia, è comunque possibile trarre delle conclusioni interessanti.

In generale possiamo sostenere che le finestre con la versione standardizzata ottengono una media di errore minore rispetto alla controparte normale, nel dettaglio, la versione FS1.STD primeggia costantemente. Lo si può osservare nel grafico 4.5 che riassume gli esiti dei problemi disegnati, esponendo l'errore medio ottenuto trasversalmente alle features. È chiaro che il valore medio in queste analisi permette di trarre una visione generale dell'insieme.

Tuttavia, si sottolinea che nei dati ottenuti sono presenti delle features che, a causa del loro scarso risultato, sbilanciano in modo negativo l'esito. Per questo verranno comunque evidenziati i dati che spiccano rispetto al gruppo, ma che potrebbero discostarsi in alcuni casi dai quelli esposti nel grafico. Si evidenzia che, siccome i casi di classificazione ternaria (3 classi) hanno avuto un errore fuori scala rispetto alla media generale, si è deciso per comprimere una parte dell'asse delle ordinate per ottenere una migliore visione dell'insieme. A seguire saranno descritti nel dettaglio i tre gruppi elencati nel capitolo 4.4.2.

In merito ai casi di classificazione PR-2.1 (binari) che discriminano la presenza/assenza di una determinata classe, i risultati mostrano un errore medio moderato. È possibile notare che la classe V ottiene un lieve miglioramento con la finestra FS0X.NOR, diversamente dagli altri casi, dove invece vediamo sempre protagonista la versione FS1.STD.

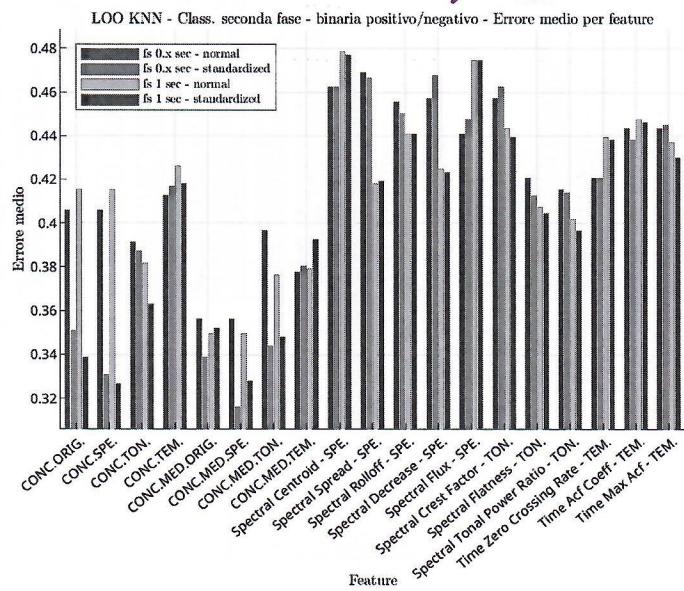
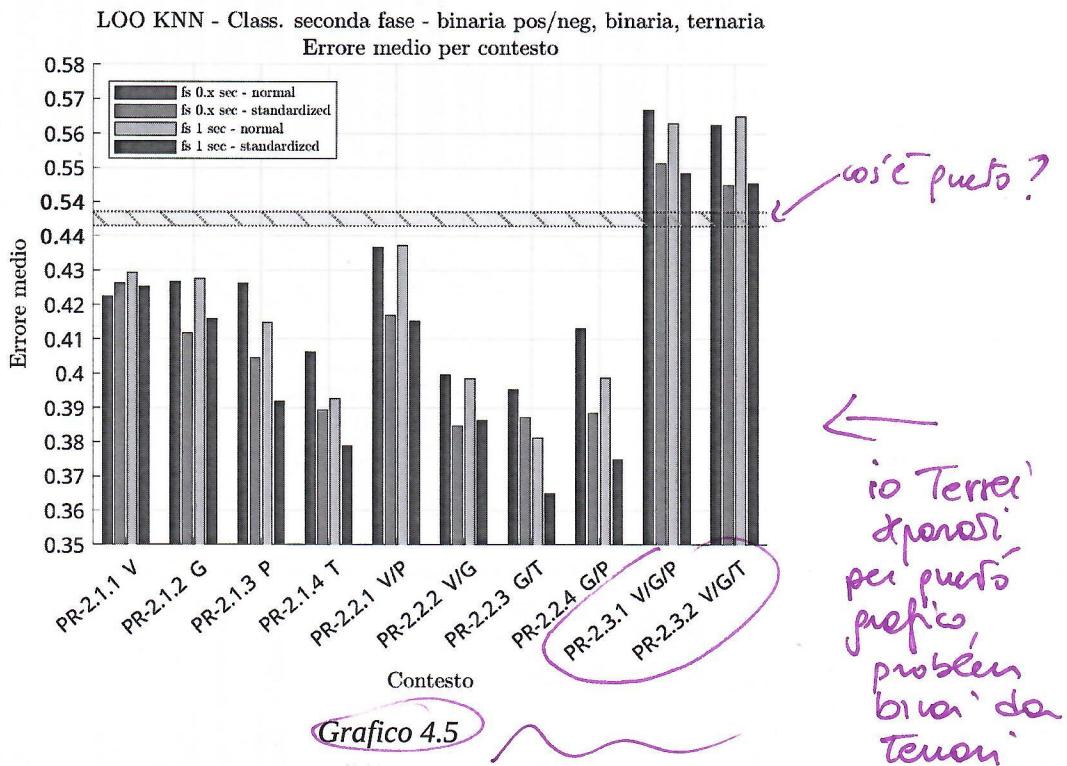
In questo gruppo i problemi con i migliori risultati riguardano i suoni appartenenti alla GEO, delle classi T e P nei problemi PR-2.1.3 P e PR-2.1.4 T. Si evidenza il risultato delle feature CONC.MED.SPE, che primeggia in entrambe le frequenze, con errori pari al 25% nel primo e 27% nel secondo, e in un posizione molto vicina anche la feature CONC.MED.ORIG.. Il loro risultato lo si può notare nel grafico 4.6, che mette in evidenza il comportamento delle features trasversalmente ai casi sottoposti, con una media sui valori. È possibile notare che anche CONC.SPE. e CONC.ORIG, hanno ottenuto in generale un buon risultato. In questo gruppo la BIO sembra essere stata caratterizzata dalle feature spettrali.

Per quanto concerne il gruppo PR-2.2, classificazione binaria ma con due classi distinte, su tre dei quattro problemi, i risultati sono soddisfacenti per il nostro contesto. Le classi si trovano in una condizione molto più bilanciata e il sistema ne ha goduto, seppur lievemente. Questi problemi hanno permesso di valutare gli insiemi ANT/BIO/GEO in contrapposizione a coppie, evidenziando delle differenze rilevanti. È possibile notare l'ANT in contrasto con la GEO, nel problema PR-2.2.1 V/P, dove ottiene risultati molto peggiori rispetto al confronto con la BIO, nel problema PR-2.2.2. V/G: è ipotizzabile supporre che il suono dei grilli, essendo discontinuo come il suono del veicolo, ha permesso al sistema di poterli discriminare più facilmente, a differenza invece, del suono della pioggia, che si mantiene persistente. Allo stesso modo, lo ritroviamo nella correlazione di BIO con ANT, in PR-2.2.1 V/P, e BIO con GEO, in PR-2.2.4 G/P: si nota nel primo una maggiore difficoltà, come appena evidenziato. La minima presenza del suono di veicoli all'interno del dato ha sicuramente influito sul modello, quindi è possibile ipotizzare che il suono dei grilli con la pioggia ~~abbia~~ ha permesso un confronto equo, essendo entrambi molto percepibili. Differentemente, potendo valutare due casi relativi allo stesso confronto BIO/GEO, in PR-2.2.3 G/T e PR-2.2.4 G/P, è osservabile un comportamento simile, con valori leggermente migliori dei casi sopra esposti. Si ipotizza che l'origine simile, per T e P, pur essendo suoni diversi, converga in un intervallo di frequenze simili, che rispecchiano la GEO. Da un punto di vista delle features, si evidenzia che, in maniera simile al primo gruppo i risultati migliori sono dati dalla versione standardizzata per CONC.ORIG e CONC.SPE per quanto riguarda il problema PR-2.2.4 G/P, con un 20% medio di errore. Invece con un 25% negli altri casi per le stesse versioni ma con le medie concatenate in CONC.MED.ORIG e CONC.MED.SPE.. Tali valori si possono notare nel grafico 4.6, che espone i risultati dalla prospettiva delle features.

Nei casi ternari, PR-2.3.1 V/G/P e PR-2.3.2 V/G/T, i risultati si sono rivelati più deludenti delle aspettative. I suddetti problemi, pur avendo classi con dati bilanciati, si trovavano in quantità significativamente minori. Inoltre, il numero di classi, pari a tre, ha complicato ulteriormente il modello, unitamente ai possibili disturbi causati dagli altri suoni non inerenti al problema. I valori si discostano di almeno il 15% dalle altre casistiche, collocandosi nell'intorno del 50esimo percentile medio di errore. Dal punto di vista delle features invece si nota un comportamento diverso dai casi precedenti. Infatti, nel primo caso PR-2.3.1 V/G/P otteniamo i migliori risultati con le feature CONC.MED.ORIG, nella versione standardizzata in particolare, invece nel secondo caso PR-2.3.2 V/G/T, vediamo protagonista la CONC.MED.TON con un errore del 40% circa, per entrambe le finestre ed entrambe le

forme normali e standardizzate. Il grafico 4.8 ci propone i dati del gruppo dal punto di vista delle features.

In sintesi, possiamo evidenziare che i tutti i casi analizzati le *feature* CONC.MED.ORIG si dimostrano tra le migliori per descrivere i vari casi esposti, seguite da CONC.MED.SPE. nel particolare dei casi binari, e da CONC.MED.TON. per i casi ternari. In particolare, esprimono al meglio il loro potenziale nella versione con la finestra *FS1.STD*, che è risultata superiore nella maggioranza delle casistiche.



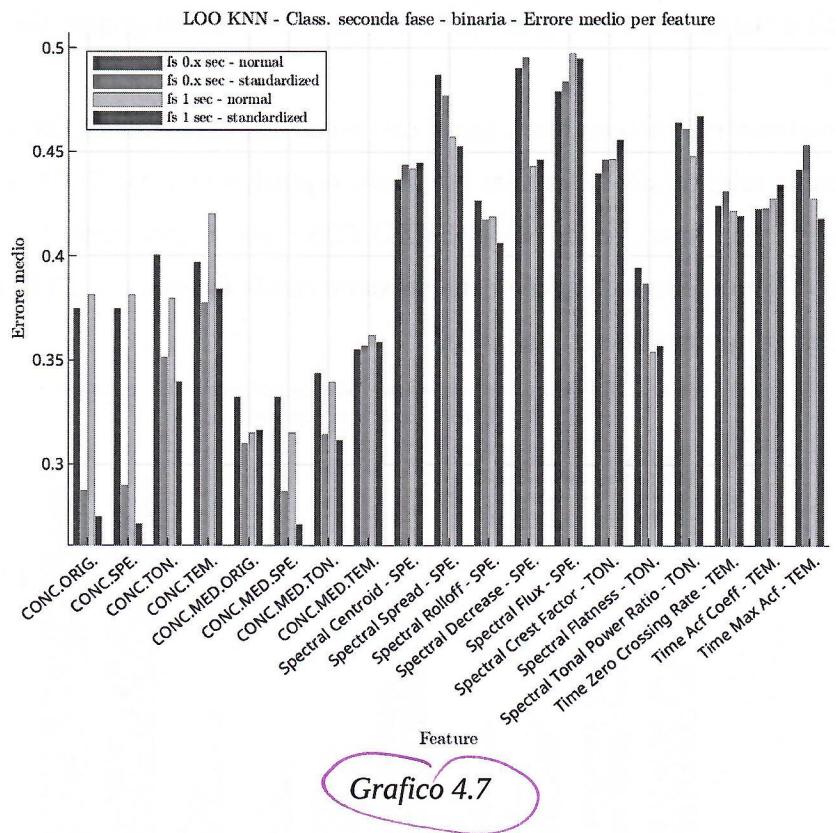


Grafico 4.7

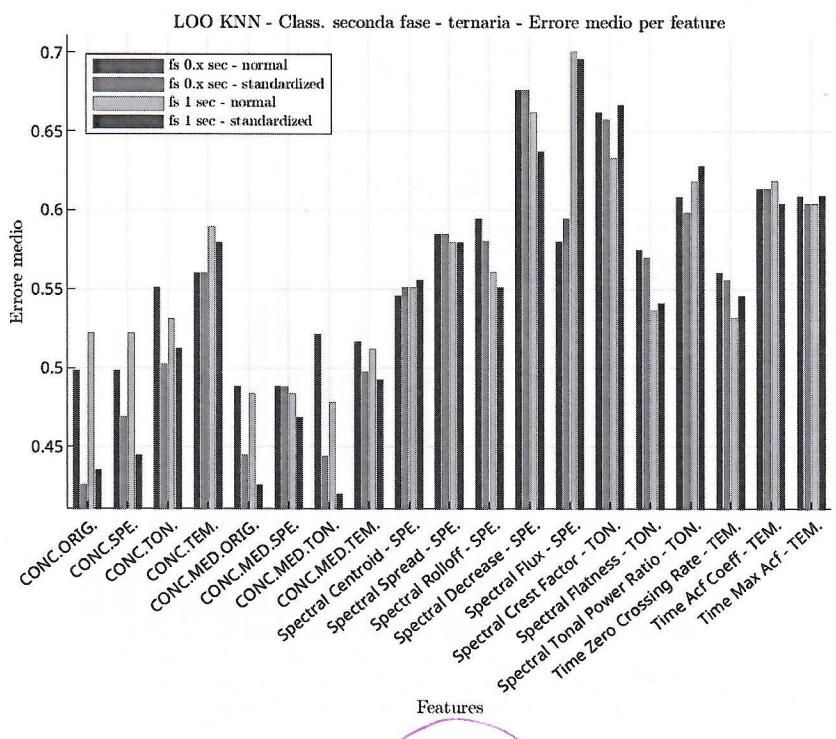


Grafico 4.8

# Capitolo 5

## Anomaly Detection

In questo capitolo si andrà ad esporre la terza fase dello studio, che si occupa di applicare algoritmi di A.D. per individuare eventuali oggetti anomali. Nel contesto dei *soundscape* gioca un ruolo molto interessante, per identificare suoni o eventi insoliti che potrebbero indicare cambiamenti ambientali, rumori estranei o fenomeni singolari: per esempio, la presenza di un cacciatore, oppure un animale non appartenente alla fauna del luogo. Nella *noite* dell'analisi, si è deciso di sperimentare l'utilizzo di questa metodologia applicando le conoscenze ottenute dalle fasi precedenti, per inferire informazioni ed esaminare quali elementi vengono classificati come anomali.

### 5.1 Configurazione *features* e metodi

Nel seguente paragrafo saranno descritte le *features* utilizzate in questo studio e le configurazioni dei metodi esposti nel paragrafo 2.3. Come menzionato sopra, la conoscenza ottenuta con lo studio di classificazione ci ha permesso di individuare *feature* che riescono ad *castigare* illustrare chiaramente il *dataset*. Inoltre, dai risultati si è potuto verificare quale delle due finestre sia in grado di definire le caratteristiche più rilevanti. Si è quindi utilizzato il set di *feature* della versione FS1 di entrambe le forme di dati, per un totale di 38 *features*, 19 normali (FS1.NOR) e 19 standardizzate (FS1.STD).

Per quanto riguarda i metodi, le versioni utilizzate presentano delle configurazioni considerabili come standard, quindi non sono state fatte particolari ottimizzazioni. IF utilizza 100 alberi decisionali, un valore largamente utilizzato. La scelta di un numero elevato determina una maggiore robustezza del modello, ma incrementa notevolmente anche il tempo di calcolo. Il numero di campionamenti presi per ogni albero è di 256 elementi, un valore comune per mantenere una buona prestazione. LOF, invece, che discrimina un'anomalia basandosi sulla densità locale definita dai *k* vicini, si è tenuto per questi ultimi il valore di venti elementi, e come metrica di distanza, quella euclidea, già descritta precedentemente.

Per l'algoritmo OCSVM, utilizza come kernel il modello gaussiano, così da proiettare i dati in uno spazio superiore dove poterli separare. Infine, come fattore di contaminazione, ovvero la soglia decisionale per definire quali elementi trattare come anomalie, è stata impostata a zero per tutti e tre gli algoritmi. Con tale valore il sistema tratta tutti gli elementi come normali e

*forse do spiegare pochino meglio*

} non come anomalie, assegnandogli solo un punteggio di anomalità. Si è considerato questo valore poiché lo studio non ha valutato tale soglia, ma solo le posizioni massime di punteggio ottenute.

## 5.2 Validazione dei risultati

In questo paragrafo verrà esposta la modalità eseguita per applicare A.D. nel contesto dello studio. Gli algoritmi sopra descritti fanno parte della tipologia di approcci non supervisionati, ovvero non necessitano quindi di etichette per poter funzionare. Però per poter dedurre informazioni rilevanti, è stato comunque impiegato il dataset DATA2, che è etichettato, così da potersi servire delle conoscenze sul contenuto per verificare il risultato.

La modalità del processo prevede 30 esecuzioni dell'algoritmo, alimentato con i dati di una feature. Dai punteggi di anomalia ottenuti viene calcolata la media, in maniera tale da essere indipendenti dal valore della singola esecuzione. L'approccio viene replicato per ciascuna feature, e per tutti gli algoritmi. Infine, da ogni esecuzione si sono considerati solo i primi tre valori in classifica con il punteggio maggiore: quindi 38 (features), per tre risultati, e per tre algoritmi. Per valutare ogni metodo, sono state estratte le prime cinque occorrenze più numerose da ognuno. Per avere visione complessiva, invece, sono state estratte le prime cinque occorrenze migliori di tutti e tre i set di risultati.

## 5.3 Risultati

Dopo aver esposto la modalità di applicazione dell'analisi si andrà a descrivere i risultati ottenuti. Nella tabella 5.1 è possibile avere una visione dei risultati come descritti nel paragrafo precedente: le maggiori occorrenze per i tre algoritmi, e poi sul totale, con le relative percentuali di distribuzione.

→ devi spiegare + bene

Dal punto di vista dei singoli suoni, vediamo come il rumore dei veicoli è stato rilevato per un 40% su IF e O, per un 60% invece per L. Il suono degli uccelli U è sempre presente per un 60% e con una distribuzione simile in tutte e tre le casistiche, diversamente invece, da quello dei grilli G, che è emerso presente per un 40% e solo in L. Il suono del fiume/cascata C, come prevedibile, appare molto diffuso tra i risultati, ma si può ipotizzare che non abbia nulla di particolare e che la sua presenza sia dovuta unicamente alla sua ampia distribuzione nei dati.

Anche il suono della pioggia si è notevolmente osservato, per un 80%, e si può affermare che valgono le stesse conclusioni appena fatte per il caso precedente. Invece, le interferenze e i suoni sconosciuti non hanno realizzato le aspettative. Si poteva prevedere che sarebbero

suo strettamente con-

nessuali, contrariamente alle aspettative

emersi data la loro minima presenza, quanto alla loro particolarità, ma invece sono stati rilevati solo per un caso, e solo da O.

Come è possibile notare, dalla prospettiva dei metodi, si vede che IF e O hanno espresso delle distribuzioni simili, diversamente da L. che ha dimostrato una sensibilità diversa nella maggioranza dei casi. (spiegare)

Invece, se si osserva dal punto di vista degli insiemi di suoni ANT/BIO/GEO, si può notare come la GEO sia molto presente, per almeno un 80%. Per una parte si possono trarre le stesse conclusioni definite prima sulla distribuzione dei suoni C e P, ma per quanto riguarda quello di T, si può ipotizzare che non lo sia. Infatti, la sua distribuzione nel dataset di riferimento è di circa del 60%, e in rapporto con la sua bilanciata presenza tra i risultati si può ritenere che l'informazione è molto interessante. Similmente, per ANT e BIO è possibile sostenere l'ipotesi che il risultato ottenuto sia rilevante: per il primo sulla bassa numerosità dei dati in input, per il secondo invece, data la sua grande distribuzione, perché non si è propagato similmente a quanto osservato nei suoni più rilevati della GEO.

Infine, da una prospettiva temporale, si manifesta che l'80% dei risultati sul totale, ma anche nei singoli algoritmi, si concentra alle ore 14:00, nella seconda parte della giornata. Lo stesso per l'elemento identificato maggiormente come anomalia che risulta invece verso le ore 18:00.

spese serie  
↓ ↘ ↗

			ANT	BIO		GEO		ALTRO		
GRUPPO	AUDIO	#	V	U	G	C	P	F	I	S
TOTALE	20200323_180000.WAV	43	x	x			x			
	20200303_140000.WAV	25					x	x		
	20200304_140000.WAV	21				x	x	x		
	20200311_140000.WAV	18				x	x	x		
	20200318_140000.WAV	14	x	x		x				
IFOREST		40%	40%	0%	80%	80%	60%	0%	0%	↙
	20200323_180000.WAV	22	x	x		x	x			
	20200303_140000.WAV	18					x	x		
	20200304_140000.WAV	12				x	x	x		
	20200309_060000.WAV	8	x	x		x				
LOF	20200316_220000.WAV	8		x		x	x			
		40%	60%	0%	80%	80%	40%	0%	0%	
	20200311_140000.WAV	15				x	x	x		
	20200323_180000.WAV	15	x	x		x	x			
	20200318_140000.WAV	10	x	x		x				
OCSVM	20200301_180000.WAV	9		x	x	x		x		
	20200325_020000.WAV	8	x		x	x				
		60%	60%	40%	100%	40%	40%	0%	0%	
	20200323_180000.WAV	6	x	x		x	x			
	20200309_060000.WAV	5	x	x		x				
	20200314_180000.WAV	4	x	x		x	x	x		x
	20200328_140000.WAV	4				x	x	x		
	20200303_140000.WAV	3					x	x		
		60%	60%	0%	80%	80%	80%	0%	0%	20%

Tabella 5.1

## Capitolo 6

### Conclusione

In questo studio sono state applicate tecniche di *Pattern Recognition* e *Machine Learning* per caratterizzare mediante classificazione e *anomaly detection* un *soundscape*.

Come è già noto nella letteratura sui *soundscape*, il contesto dello studio presenta notevoli difficoltà intrinseche. Si deve tenere conto che lo stesso cervello umano, nonostante le sue straordinarie capacità, fatica a distinguere e classificare determinati suoni.

Entrambe le fasi di classificazione hanno dimostrato che la *feature* migliore consiste nella concatenazione delle medie dei gruppi di tipologie di *features*, nella versione standardizzata e in combinazione con la dimensione della finestra più ampia. Tra i problemi con categorie note la discriminazione sul luogo ha ottenuto risultati nettamente superiori rispetto al resto.

Nelle casistiche con categorie semantiche i problemi binari con due classi distinte occupano la prima posizione distaccandosi di poco da quelli binari che discriminano la presenza/assenza della classe. I problemi ternari, essendo un contesto multi etichetta, hanno affrontato le maggiori difficoltà e i risultati si sono espressi con una predizione media corretta minore della metà degli oggetti. Nel particolare caso dei binari le feature spettrali hanno ottenuto risultati interessanti, per i ternari lo sono state quelle tonali.

L'*Anomaly detection* non ha rilevato particolari elementi anomali e non è emerso nessuno schema particolare tra i dati. Le aspettative riguardanti i suoni relativi a interferenze e a quelli sconosciuti non si sono rivelate; non sono emersi come anomalie. I suoni relativi alla geofonia sono stati riscontrati frequentemente, in particolare quello del tuono ha ottenuto un risultato interessante. Si è potuto notare che la presenza di dati anomali ricade nelle fasce orarie pomeridiane.

Il tentativo di filtraggio sperimentato non ha portato nessun vantaggio dal punto di vista della qualità dell'audio, ma i risultati ottenuti sono comunque validi. La stessa combinazione di caratteristiche vincenti emerse dalla classificazione si sono mantenute tali anche rimuovendo parte delle frequenze.

Noti i vantaggi derivati dalla comprensione dei *soundscape*, si termina questo studio con la speranza che la ricerca possa avanzare e trovare soluzioni migliori per approfondire la conoscenza del nostro ecosistema.

metti a fare ricerca  
abbiamo realizzato  
una ricerca, abbiamo  
sviluppato diversi  
feature...

com'è  
com'è  
va fatto  
come

occhio  
è  
esso?

## Bibliografia

- mea' leri e uso' complete
- [1] Colin A. Quinn et alii. *Soundscape classification with convolutional neural networks reveals temporal and geographic patterns in ecoacoustic data*. Ecological Indicators, Vol. 138, 2022, 108831, ISSN 1470-160X, <https://doi.org/10.1016/j.ecolind.2022.108831>
  - [2] Farina Almo et alii, *Sonotopes reveal dynamic spatio-temporal patterns in a rural landscape of Northern Italy*. Frontiers in Ecology and Evolution, vol.11 ,2023, 2296-701X, <https://www.frontiersin.org/journals/ecology-and-evolution/articles/10.3389/fevo.2023.1205272>
  - [2] Tovar Garcia, J.D. and Acevedo-Charry, O. *Dataset of passive acoustic monitoring at the Nature Reserve Los Yátaros, Gachantivá, Boyacá, Colombia*. 2021. Biota Colombiana. 22, 1 (Jan. 2021), 200–208. <https://doi.org/10.21068/c2021.v22n01a13>
  - [4] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. 2012. IEEE Press. par. 3.3.3
  - [5] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. 2012. IEEE Press. par. 3.3.4
  - [6] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. 2012. IEEE Press. par. 3.3.1
  - [7] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. 2012. IEEE Press. par. 3.3.5
  - [8] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. 2012. IEEE Press. par. 3.3.2
  - [9] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. 2012. IEEE Press. par. 3.4.1.1
  - [10] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. 2012. IEEE Press. par. 3.4.3
  - [11] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. 2012. IEEE Press. par. 5.3.3.1
  - [12] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. 2012. IEEE Press. par. 5.3.3.2
  - [13] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. 2012. IEEE Press. par. 5.3.3.6

> Baita 1, con seve indices d'informació