

# Underwater Classification

Giuseppe D'Avino, Mario Lezzi and Daniele Dello Russo

---

## ABSTRACT

Questo studio si concentra sulla distinzione tra suoni generati dagli esseri umani (Target) e quelli prodotti dai pesci (Non Target), con l'obiettivo di migliorare il monitoraggio degli ecosistemi marini e mitigare l'impatto delle attività antropogeniche. Utilizzando tecniche avanzate di machine learning e analisi del segnale, è stato sviluppato un sistema di classificazione automatica in grado di identificare e separare con precisione i suoni antropogenici dai suoni della fauna marina. La metodologia comprende la raccolta di un ampio dataset di registrazioni di sample subacquee, l'estrazione di caratteristiche distintive come spettri di frequenza e MFCC, e l'addestramento di modelli di Machine Learning. L'addestramento è stato suddiviso in due fasi principali: nella prima fase, è stato implementato una classificazione binaria, in cui il modello doveva determinare se ogni registrazione dei sample contenesse suoni generati da esseri umani o da pesci. Nella seconda fase, è stata applicata una classificazione multiclasse per i suoni generati dai pesci. In questa fase, il sistema è stato addestrato a riconoscere a quale sottoclasse specifica appartenesse ogni sample.

---

## Underwater Classification

### 1. Introduzione

Nell'era dell'informazione, la quantità di dati audio generati e raccolti è in continua crescita, con l'audio che rappresenta una fonte ricca e complessa di informazioni. Tuttavia, gestire e analizzare questi dati costituisce una sfida rilevante, soprattutto quando il volume diventa considerevole. Diversi ambiti richiedono la gestione e l'analisi di grandi quantità di dati audio: ad esempio, l'oceano è un ambiente acusticamente ricco e complesso, caratterizzato dalla continua interazione tra suoni naturali e artificiali. Con lo sviluppo di tecnologie avanzate per la registrazione subacquea, la raccolta di dati acustici marini è divenuta una pratica comune per numerosi scopi, come la ricerca scientifica, la conservazione della fauna marina e il monitoraggio delle attività umane. In questo contesto, risulta cruciale approfondire la comprensione e l'analisi dei segnali acustici marini, al fine di classificarli e distinguere tra suoni generati da attività antropiche (Target) e suoni prodotti dalla fauna marina, in particolare dai pesci (Non Target). Questa distinzione riveste un'importanza fondamentale per diverse ragioni: dalla protezione e conservazione degli ecosistemi marini, al miglioramento della ricerca scientifica, fino allo sviluppo di tecnologie innovative e strumenti di monitoraggio ambientale. L'impiego di tecniche avanzate di machine learning e di analisi del segnale acustico fornisce una soluzione promettente per una gestione più sostenibile e consapevole degli ecosistemi marini.

Gli obiettivi principali di questo studio sono i seguenti:

(i) Condurre un'analisi approfondita della letteratura esistente sul riconoscimento dei segnali acustici in ambienti marini, al fine di identificare le caratteristiche fondamentali dei segnali audio in questo contesto. Questo passaggio permetterà di creare un template delle caratteristiche

essenziali per l'identificazione dei suoni prodotti dalla fauna marina e di quelli antropogenici.

(ii) Sviluppare un modello di classificazione in grado di distinguere i suoni bioacustici marini dalle registrazioni audio, separando in modo efficace i segnali biologici da quelli di origine antropica. In seguito, l'obiettivo sarà migliorare il modello per distinguere non solo a livello intraclasse, ma anche interclasse, con lo scopo di identificare specificamente le specie animali rilevate nei segnali analizzati.

(iii) Ottimizzare le caratteristiche numeriche estratte da un segnale audio per descriverlo in maniera chiara e dettagliata. A tal proposito, saranno discusse diverse caratteristiche acustiche e il loro potenziale informativo nel contesto della ricerca scientifica e dell'identificazione dei suoni.

Questi obiettivi rappresentano un passo cruciale verso l'elaborazione di strumenti efficaci per la comprensione dei segnali acustici marini, con implicazioni significative sia per la conservazione ambientale che per il progresso scientifico.

### 2. Stato dell'arte

L'analisi bioacustica è un campo in rapida espansione, sostenuto dall'integrazione di nuove tecnologie, come l'intelligenza artificiale, che sta rivoluzionando ogni aspetto dell'analisi dei dati. Prima dell'avvento dell'IA, l'analisi bioacustica veniva eseguita mediante una combinazione di tecniche manuali e strumenti acustici tradizionali. Questi metodi erano spesso laboriosi e richiedevano un intervento umano significativo per la raccolta, l'analisi e l'interpretazione dei dati sonori. È evidente che, prima dell'introduzione delle nuove tecnologie, ogni dato bioacustico doveva essere analizzato singolarmente, trattato come caso di studio e sottoposto all'esame di esperti del settore, rallentando notevolmente le fasi di analisi e classificazione dei segnali. (6)(2)(Gibb et al., 2019; Blumstein et al., 2011). Nel contesto dell'analisi

---

ORCID(s):

e della classificazione, i passaggi fondamentali possono essere così riassunti in:

**Raccolta dati:** Questa fase veniva realizzata mediante la registrazione degli audio e il campionamento manuale, richiedendo la presenza fisica degli esperti per distinguere accuratamente i segnali bioacustici da raccogliere.

**Pre-elaborazione dei dati :** Durante questa fase, si procedeva al filtraggio e alla pulizia delle fonti per ridurre il rumore di fondo e migliorare la qualità del suono registrato, utilizzando filtri passa-basso (LowPass) e passa-alto (HighPass). La segmentazione veniva eseguita manualmente per selezionare solo gli elementi rilevanti per il contesto di applicazione.

**Analisi del suono:** I segnali raccolti venivano analizzati uno per uno, trasformandoli in spettrogrammi e oscillogrammi, analizzando la frequenza e effettuando misurazioni manuali.

**Classificazione e identificazione :** In questa fase, venivano utilizzate chiavi dicotomiche per identificare e distinguere le diverse specie animali basandosi sulle vocalizzazioni, che presentano caratteristiche specifiche per ciascuna specie. Inoltre, veniva effettuato un confronto diretto con samples noti appartenenti a determinate specie.

**Documentazione e archiviazione:** Infine, si procedeva alla catalogazione, associando le registrazioni alle varie specie animali conosciute; in caso di assenza di corrispondenza, i segnali venivano classificati come componenti bioacustiche inconsistenti o sospettati di appartenere a sottospecie già esistenti.

Le limitazioni delle analisi pre-IA includevano il tempo impiegato e le ingenti risorse necessarie per ogni caso di studio. Inoltre, vi era una componente di soggettività e interpretazione personale da parte degli esperti, nonché una capacità limitata di elaborazione, poiché gli esperti di un determinato settore potevano lavorare solo su una parte delle ricerche alla volta.(6)(2) (Gibb et al., 2019; Blumstein et al., 2011).

Oggi, con l'introduzione dell'IA, sono state sviluppate diverse soluzioni per automatizzare le fasi precedentemente descritte. Nell'ambito dell'analisi dei segnali bioacustici, si utilizza un nuovo approccio basato sull'impiego dell'IA per automatizzare e analizzare empiricamente le fonti audio. In questo contesto, sono descritti diversi progetti su cui si basa il nostro lavoro di ricerca e sviluppo di un nuovo sistema in grado di eseguire una distinzione interclasse e intraclasse dei segnali bioacustici in ambiente sottomarino.(9) (Kahl et al., 2021).

### 3. Caso di studio: BirdNet

QuEcoacoustics/audio-analysis:QUT Ecoacoustics Analysis Programs è un pacchetto software in grado di eseguire una serie di analisi su registrazioni audio ambientali. Sebbene queste analisi siano progettate per registrazioni di lunga durata (1-24 ore), possono essere eseguite su qualsiasi sample in un formato supportato dal software. Il software è capace di:

- Calcolare indici acustici spettrali e sommari a risoluzioni variabili.
- Produrre spettrogrammi multi-indice, falsi colori e di lunga durata.
- Calcolare statistiche critiche di annotazioni scaricate da un Acoustic Workbench.
- Eseguire vari riconoscitori di eventi acustici.

Sebbene questo progetto non sia in grado di distinguere tra segnali bioacustici e non, fornisce un'importante linea guida su come strutturare l'analisi di sample relativi a questa specifica classe. Il modello di AI utilizzato è un classificatore lineare basato su reti neurali. In particolare, il codice permette di costruire un classificatore lineare con uno o due strati nascosti, controllato dal parametro `hidden_units`. Il modello è implementato utilizzando la libreria Keras (importata attraverso `keras`), e il processo di addestramento è potenziato da un algoritmo di ottimizzazione bayesiana per il tuning automatico degli iperparametri (utilizzando `keras_tuner.BayesianOptimization`). Il nostro obiettivo è stato quello di utilizzare le conoscenze acquisite da questo progetto come base di partenza per il nostro applicativo. La limitazione fondamentale di questo progetto è sicuramente l'utilizzo di registrazione di lunga durata per poter distinguere in maniera chiara i segnali bioacustici, inoltre il dataset utilizzato per l'analisi risulta essere verticali andando a considerare solo volatili, infine l'applicativo non tiene conto della distinzione dei suoni antropogenici. Successivamente al progetto BirdNET, che ha evidenziato l'efficacia dell'IA nel riconoscimento dei canti degli uccelli, è emerso un approccio innovativo nell'ambito dell'analisi bioacustica applicata al monitoraggio marino. L'articolo (11) (Malfante et al 2018), presenta un avanzato metodo per il monitoraggio passivo della vitalità degli oceani, con un focus particolare sulle popolazioni di pesci. Lo studio sviluppa un modello discriminativo basato su tecniche di machine learning supervisionato, in particolare Random Forest (RF) e Support Vector Machines (SVM), per classificare i suoni dei pesci. Il modello si distingue per l'uso di caratteristiche estratte dai domini temporale, frequenziale e cepstrale dei segnali acustici. Testato su suoni reali registrati in diverse aree marine, il sistema ha raggiunto un'accuratezza di classificazione del 96,9. Un aspetto particolarmente rilevante dello studio è l'approccio dettagliato all'estrazione delle caratteristiche. Gli autori esplorano l'impatto delle caratteristiche provenienti dai diversi domini, dimostrando che sebbene ciascun dominio contenga informazioni discriminative utili, la combinazione delle caratteristiche estratte dai tre domini offre prestazioni superiori. Per ottimizzare la classificazione, è stato adottato un metodo di selezione delle caratteristiche basato sui pesi delle caratteristiche nel modello RF. Sono stati identificati due sottoinsiemi di caratteristiche: Most Valuable Features (MVF) e Valuable Features (VF). Il MVF comprende tre caratteristiche

chiave—l'energia kurtosis dal dominio frequenziale (F28), la kurtosis media dal dominio temporale (T7) e il tasso di attraversamento della soglia dal dominio temporale (T15)—e ha raggiunto un'accuratezza media del 91,5% (RF) e del 91,3% (SVM). Il set VF, che include il MVF e altre 16 caratteristiche aggiuntive, ha migliorato l'accuratezza globale al 95,6% (RF) e al 94,7% (SVM). Questi risultati suggeriscono che, sebbene l'uso di tutte le caratteristiche possa non essere necessario per ottenere risultati di alta qualità, la selezione mirata delle caratteristiche consente di mantenere l'efficacia del sistema di classificazione. Questo è particolarmente utile per applicazioni in tempo reale con risorse computazionali limitate. Inoltre, l'analisi evidenzia che l'importanza delle caratteristiche può variare a seconda della classe, sottolineando la necessità di una selezione accurata per ottimizzare la classificazione in base ai requisiti specifici delle diverse caratteristiche. (11) (Malfante et al. 2018) In sintesi, l'integrazione dell'intelligenza artificiale ha rivoluzionato l'analisi bioacustica, migliorando l'efficienza e la precisione nella classificazione dei segnali sonori. Progetti come BirdNET e l'approccio descritto da Malfante et al. dimostrano che l'uso combinato di tecniche avanzate e selezione mirata delle caratteristiche consente risultati superiori nella rilevazione e classificazione dei suoni, sia in contesti terrestri che marini. Questi progressi evidenziano il potenziale dell'IA nel trasformare l'analisi bioacustica di suoni.

#### 4. Struttura del Dataset

Il dataset utilizzato in questo studio è stato suddiviso in due principali categorie: Target e Non Target (96 sotto-directory). La directory Target (16 sotto-directory) include segnali audio di origine antropogenica, mentre la directory Non Target contiene segnali BioAcustici. Le registrazioni per la creazione di questo dataset sono state estratte da fonti disponibili online, con l'obiettivo di costruire un dataset composito da utilizzare come base per lo sviluppo del modello di intelligenza artificiale. Tuttavia, è nostra intenzione, in futuro, utilizzare un dataset creato ad hoc per migliorare la qualità e la precisione del modello. Essendo un dataset eterogeneo, con un numero di samples di 2663, le caratteristiche tecniche dei segnali audio variano considerevolmente: il campionamento presenta un range di frequenza che va da 600 Hz a 384.000 Hz, con segnali sia in formato mono che stereo. L'ampiezza del segnale varia da 0 a 2, mentre la profondità in bit (bit depth) spazia da 8 bit PCM fino a 32 bit. La durata dei segnali audio è anch'essa estremamente variabile, con registrazioni che durano da pochi secondi fino a oltre 30 minuti. Inoltre, il dataset risulta sbilanciato tra le categorie di segnali. Le manipolazioni e gli interventi necessari per bilanciare e preparare i dati saranno descritti in dettaglio nelle sezioni successive.

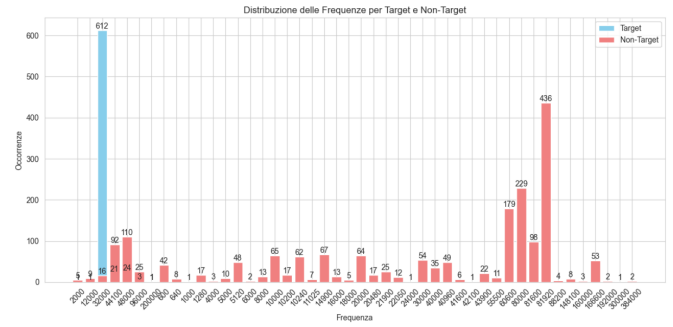


Figure 1: Distribuzione frequenza target e non target

#### 5. Analisi

Il primo passo del nostro studio ha riguardato un'accurata analisi preliminare del dataset. In questa fase, è stata condotta una ricerca approfondita e la rimozione dei file duplicati, garantendo l'unicità di ciascun segnale audio. Successivamente, per ogni sample, è stato calcolato e registrato le seguenti proprietà fondamentali:

- **Ampiezza** : utilizzata per valutare l'intensità del segnale audio, al fine di comprendere meglio le dinamiche sonore presenti nelle registrazioni.
- **Numero di canali**: impiegato per determinare se i segnali audio fossero in formato mono o stereo.
- **Bit Depth**: indicante la risoluzione dei dati audio, ovvero la precisione con cui è rappresentato ogni sample del segnale.
- **Frequenza di campionamento**: rappresentante il numero di samples acquisiti al secondo, utilizzata per valutare la qualità temporale del segnale.
- **Durata**: misurata in termini di lunghezza temporale di ciascun segnale audio, poiché essa varia significativamente all'interno del dataset.
- **Fase**: si riferisce alla posizione relativa del punto di partenza di un'oscillazione o vibrazione rispetto a un riferimento temporale. Essa indica in che punto del ciclo l'onda si trova in un determinato istante.

Questi parametri ci hanno fornito una panoramica dettagliata delle caratteristiche del dataset e hanno costituito la base per le successive fasi di pre-elaborazione dei dati.

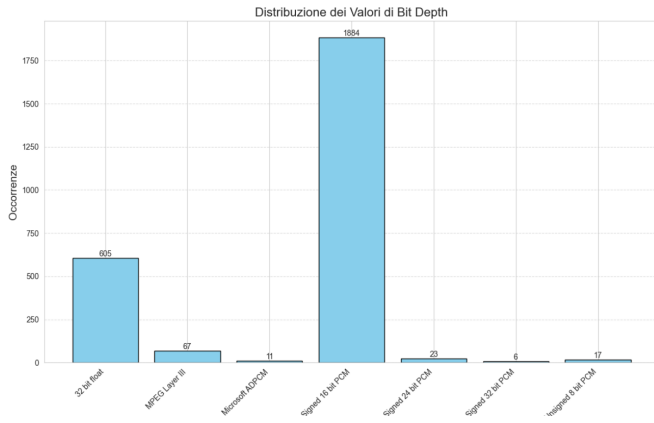


Figure 2: Distribuzione valori Bit Depth in Target

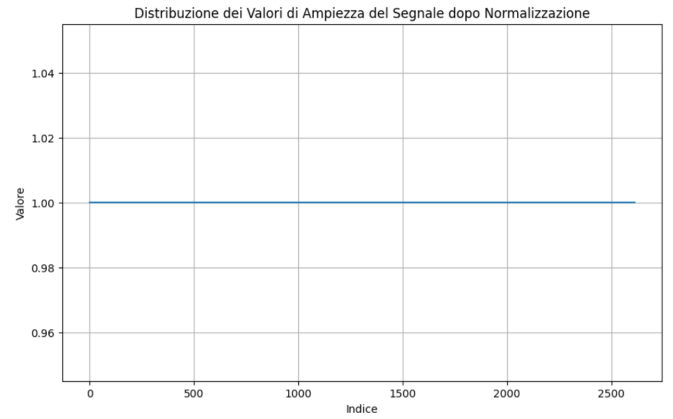


Figure 4: Ampiezza segnali dopo la normalizzazione

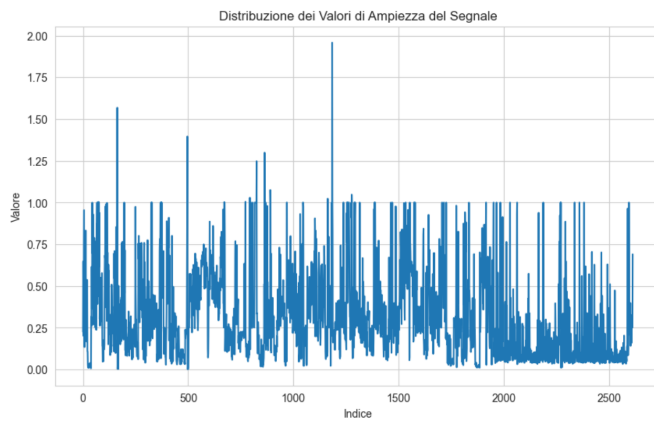


Figure 3: Ampiezza segnali

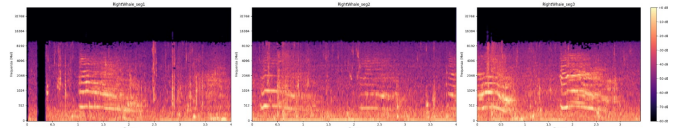


Figure 5: Spettrogrammi sample audio segmentato

## 6. Pre-Processing

Dopo aver completato l'analisi preliminare, è stato eseguito il pre-processing dei dati audio per prepararli all'addestramento del modello. In primo luogo, i segnali audio sono stati normalizzati per uniformare l'ampiezza (compresa tra -1 e 1) e ridurre le variazioni estreme che avrebbero potuto compromettere le prestazioni del modello. La normalizzazione ha reso i segnali più omogenei e comparabili. Successivamente, tutte le registrazioni sono state convertite a una profondità di bit di 16-bit in monocanale per standardizzare la qualità audio, garantendo una risoluzione uniforme su tutti i dati e facilitando il processo di apprendimento del modello. Inoltre, la frequenza di campionamento di tutti i samples è stata uniformata a un valore standard di 96 kHz. Questo passaggio era fondamentale per assicurare la confrontabilità tra le diverse registrazioni e garantire che il modello ricevesse input coerenti. Le registrazioni sono state poi suddivise in segmenti di durata uniforme di 4 secondi, un'operazione che semplifica l'elaborazione e l'analisi dei dati. Nel caso di registrazioni più brevi, sono stati aggiunti periodi di silenzio per raggiungere la lunghezza desiderata. Il risultato prodotto è un dataset contenente 50993 samples.

Questi passaggi di pre-processing sono stati essenziali per garantire che i dati fossero coerenti e di alta qualità, ponendo solide basi per l'addestramento efficace dei modelli di machine learning.

## 7. Metodologie

L'estrazione delle features dai segnali audio rappresenta un passo fondamentale nella costruzione di modelli di intelligenza artificiale per l'analisi bioacustica. Ogni feature numerica selezionata consente di descrivere specifiche proprietà del segnale, facilitando la distinzione tra suoni bioacustici e antropogenici, nonché la classificazione delle sorgenti sonore. In questo contesto, le features spetro-temporali forniscono informazioni dettagliate sul contenuto spettrale e sulla struttura temporale dei segnali, che sono fondamentali per una corretta identificazione e classificazione.

- **Spectral Centroid Mean** Il centroide spettrale rappresenta la media ponderata delle frequenze presenti in un segnale, e può essere considerato come un indicatore della "luminosità" del suono. In termini matematici, il centroide spettrale è definito come:

$$\frac{1}{E} \sum_i i \cdot E_i$$

dove  $i$  rappresenta la posizione o l'indice di un elemento (ad esempio, un frame o un bin di frequenza),  $E_i$  rappresenta l'energia associata a quella particolare posizione  $i$ ,  $E$  è l'energia totale del segnale, calcolata come somma di tutte le  $E_i$ . Nelle applicazioni di segnale e audio, questa può essere vista come una misura di dove si trova il "baricentro

energetico" in un determinato dominio (come il dominio delle frequenze, ad esempio nello spettro audio).

- **Spectral Bandwidth RMS** La larghezza di banda spettrale RMS misura la dispersione delle frequenze attorno al centroide spettrale e riflette la complessità del segnale:

$$RMS_i = \sqrt{\frac{1}{E} \sum_i i^2 E_i - i^{-2}}$$

Questa feature è cruciale per distinguere tra suoni semplici e complessi, dove i suoni antropogenici tendono ad avere una maggiore larghezza di banda rispetto ai segnali bioacustici (10)(Lerch, 2012).

- **Standard Deviation** La deviazione standard dello spettro descrive la variazione delle frequenze rispetto al centroide spettrale:

$$\sigma_s = \sqrt{\frac{1}{n-1} \sum_i (s[i] - \mu_s)^2}$$

Questa feature permette di valutare la stabilità spettrale del segnale, utile per identificare suoni con fluttuazioni frequenti tipiche di alcuni rumori antropogenici(16) (Sharma et al., 2020).

- **Skewness** L'asimmetria (Skewness) dello spettro quantifica la simmetria della distribuzione delle frequenze attorno al centroide:

$$\frac{1}{n} \sum_i \left( \frac{s[i] - \mu_s}{\sigma_s} \right)^3$$

Valori positivi indicano una prevalenza di frequenze più alte rispetto alla media, mentre valori negativi indicano il contrario. Questa feature è essenziale per distinguere suoni con distribuzioni spettrali atipiche, come quelli prodotti da alcune specie marine (1)(Bishop, 2006).

- **Kurtosis** La curtosi misura la "puntosità" della distribuzione delle frequenze:

$$\frac{1}{n} \sum_i \left( \frac{s[i] - \mu_s}{\sigma_s} \right)^4$$

Una curtosi elevata indica la presenza di picchi spettrali accentuati, tipici di suoni impulsivi come i clic dei cetacei, utili per identificare eventi sonori specifici (5)(Figuerola et al., 2015).

- **Mean skewness** La mean skewness è la media della skewness calcolata su più segmenti o finestre del segnale, come nelle serie temporali o nei dati audio. Si divide il segnale in più parti (finestre) e si calcola

la skewness per ciascuna finestra, ottenendo infine la media di questi valori. La formula è la seguente:

$$\sqrt{\frac{\sum_i (i - \bar{i})^3 \cdot E_i}{E \cdot \text{rms}_i^3}}$$

Nel caso dei segnali audio, la mean skewness può essere utilizzata per capire se il segnale ha una tendenza a concentrare più energia su un lato della distribuzione, come nelle frequenze più alte o più basse, e come questa tendenza cambia nel tempo. Una skewness media potrebbe fornire un'indicazione generale della distribuzione dell'energia nel segnale.

- **Mean kurtosis** La Mean Kurtosis è una misura statistica che indica la "piccolezza" della distribuzione di un segnale, valutando quanto i dati siano concentrati attorno alla media. Nella pratica, valori elevati di kurtosi possono segnalare eventi rari o difetti nel segnale, mentre valori bassi indicano una distribuzione più uniforme.

$$\sqrt{\frac{\sum_i (i - \bar{i})^4 \cdot E_i}{E \cdot \text{rms}_i^4}}$$

- **Shannon Entropy** L'entropia di Shannon misura la quantità di informazione o complessità del segnale:

$$-\sum_j p(s_j) \log_2(p(s_j))$$

Questa feature è indicativa della complessità del segnale, con valori più alti che suggeriscono una maggiore variabilità, utile per differenziare tra suoni complessi e semplici(4)(Cover & Thomas, 2006).

- **Renyi Entropy** l'entropia di Renyi permette di analizzare segnali che presentano una distribuzione di probabilità complessa o non uniforme, il che è comune in molte applicazioni, come la compressione dei segnali e il riconoscimento dei modelli. A differenza dell'entropia di Shannon, che considera solo le probabilità degli eventi, l'entropia di Renyi introduce un parametro ( $\alpha$ ) che consente di pesare in modo differente gli eventi. L'entropia di Renyi è definita come segue:

$$\frac{1}{1-\alpha} \log_2 \left( \sum_j p(s_j)^\alpha \right)$$

Utilizzando l'entropia di Renyi, si possono confrontare segnali di diversa complessità e determinare la loro informazione contenuta. Ad esempio, valori elevati di  $\alpha$  enfatizzano eventi rari, mentre valori più bassi si concentrano su eventi più probabili, rendendo questa misura utile per applicazioni in cui è necessario rilevare anomalie o variazioni significative nel segnale.



- **Rate of Attack** Il rate of attack è una misura che quantifica la velocità con cui un segnale aumenta in ampiezza in un intervallo di tempo specifico. Nella progettazione di sistemi audio e nel trattamento dei segnali, questo parametro è particolarmente importante per comprendere come un segnale o un suono, evolve nel tempo.

$$\max_i \left( \frac{s[i] - s[i-1]}{n} \right)$$

- **Rate of decay** Il Rate of Decay è un parametro importante nell'analisi dei segnali audio e della loro percezione, che indica la velocità con cui un suono perde la sua intensità dopo aver raggiunto il picco.

$$\min \frac{s[i] - s[i+1]}{n}$$

Quando un suono viene prodotto, la sua intensità iniziale è massima. Tuttavia, nel tempo, la pressione sonora inizia a diminuire a causa di vari fattori, come l'assorbimento del suono nell'ambiente e la dispersione. Il Rate of Decay descrive quanto rapidamente avviene questo processo.

- **Silence Ratio** Il silence ratio quantifica la proporzione di "silenzio" nel segnale, definito come la percentuale di frame al di sotto di una certa soglia:

$$\frac{\#(s \text{ where } s < \text{threshold})}{\sum_i^n s[i]}$$

Questa feature è utile per discriminare tra suoni continui e suoni intermittenti, spesso associati a fenomeni naturali e antropogenici rispettivamente (7)(Giannakopoulos & Pikrakis, 2014).

- **Threshold Crossing Rate** Il Threshold Crossing Rate misura la frequenza con cui il segnale supera una determinata soglia di ampiezza:

$$\frac{\#(\text{Threshold Crossing})}{n}$$

Questa feature è utile per identificare eventi sonori distinti come clic o picchi, frequenti in suoni bioacustici come quelli emessi dai mammiferi marini (14)(Popescu et al., 2009).

- **Mean** La media fornisce un indicatore del livello generale del segnale.

$$\frac{1}{n} \sum_{i=1}^n s[i]$$

- **Max over mean** La Max over Mean è una misura statistica utilizzata nell'analisi dei segnali per descrivere la relazione tra il valore massimo di un

segnale e la sua media. Questa misura è particolarmente utile per valutare la variabilità e la distribuzione dei valori di un segnale nel tempo. La formula matematica è così definita:

$$\text{Max over Mean} = \frac{\max(s[i])}{\frac{1}{n} \sum_i s[i]}$$

- **Min over mean** La Min over Mean è una misura statistica utilizzata nell'analisi dei segnali e dell'audio. Essa rappresenta il rapporto tra il valore minimo di un segnale e la sua media. Questa misura può fornire informazioni utili sulla distribuzione dei valori nel segnale e sulle sue caratteristiche generali. La formula per calcolare la Min over Mean è:

$$\text{Min over Mean} = \frac{\min(s[i])}{\frac{1}{n} \sum_i s[i]}$$

- **Energy measurements** Le Energy measurements (misurazioni dell'energia) si riferiscono a tecniche utilizzate per quantificare l'energia contenuta in un segnale audio o in un'onda sonora. La formula per calcolare le Energy measurements è:

### Energia per un segnale discreto

$$E = \sum_{n=0}^{N-1} |x(n)|^2$$

### Energia per un segnale continuo

$$E = \int_{-\infty}^{\infty} |x(t)|^2 dt$$

- **MFCC: Mel-Frequency Cepstral Coefficients** Questi coefficienti sono ampiamente utilizzati nell'analisi audio per rappresentare la forma d'onda in modo che si adatti meglio alla percezione umana delle frequenze. I MFCC catturano le caratteristiche spettrali del segnale, rendendoli particolarmente utili per identificare timbri e texture sonore. La formula per calcolare il n-esimo MFCC è:

$$c_n = \sum_{k=1}^K \log(S(k)) \cdot \cos\left(\frac{n\pi(k - \frac{1}{2})}{K}\right)$$

- **ZCR: Zero-Crossing Rate** Questa misura indica la frequenza con cui un segnale attraversa lo zero, fornendo informazioni utili sulla complessità temporale del suono. Viene calcolata come:

$$ZCR = \frac{1}{N} \sum_{n=1}^{N-1} |\text{sgn}(x[n]) - \text{sgn}(x[n-1])|$$

- **Spectral Centroid** Questa caratteristica rappresenta il centro di massa dello spettro di potenza, indicativo della tonalità predominante nel segnale audio. La formula per calcolarlo è:

$$C = \frac{\sum_f f \cdot |X(f)|^2}{\sum_f |X(f)|^2}$$

- **Spectral Bandwidth** Questa caratteristica misura la dispersione delle frequenze rispetto al centroid spettrale e si calcola come:

$$BW = \frac{\sum_f (f - C)^2 \cdot |X(f)|^2}{\sum_f |X(f)|^2}$$

- **Chroma Features** Le Chroma Features rappresentano le intensità delle dodici note musicali suonate in un brano, permettendo di analizzare l'armonia e la tonalità. Queste caratteristiche sono particolarmente preziose per comprendere le relazioni armoniche tra le note nel contesto dei suoni marini, contribuendo così a un'analisi più ricca e stratificata del dataset. La formula è:

$$C_k = \sum_f |X(f)|^2 \cdot h(f, k)$$

- **Spectral Contrast** Il contrasto spettrale è una misura che descrive le differenze di intensità tra le bande di frequenza in uno spettro audio. Esso cattura l'interazione tra frequenze adiacenti e fornisce una rappresentazione della struttura armonica di un segnale. Si calcola con la seguente formula:

$$SC_k = \frac{1}{N} \sum_{n=1}^N \max(0, |X(f_n)|^2 - |X(f_k)|^2)$$

## 8. Oversampling & Classificazione

Inizialmente, è stata condotta una classificazione binaria per distinguere tra suoni antropogenici e bioacustici. Questa prima fase ha consentito di sviluppare un modello semplice, in grado di identificare con buona accuratezza i due gruppi principali di suoni. Successivamente, il problema è stato esteso a una classificazione multiclasse, con l'obiettivo di categorizzare ulteriormente i suoni bioacustici in diverse sottoclassi, come specie o tipi di vocalizzazioni. Questo approccio ha fornito una comprensione più profonda e dettagliata delle caratteristiche audio presenti nel dataset. Per addestrare e testare il modello, il dataset è stato suddiviso in tre parti: l'80% dei dati è stato dedicato al training, il 10% alla validazione e il 10% al test. Questa divisione ha garantito che il modello fosse addestrato su una porzione sufficientemente ampia di dati, mantenendo al

contempo un set indipendente per valutare le prestazioni durante l'addestramento e per testare la generalizzazione su dati mai visti.

Prima dell'applicazione di SMOTE, il set di addestramento per la classificazione binaria contava 34.635 samples per la classe Target e 5.791 per la classe Non-Target. Dopo l'uso di SMOTE, entrambe le classi sono state bilanciate, portando il numero di samples a 9.181 per ciascuna sottoclasse Target e 1.087 per ciascuna sottoclasse Non-Target. In seguito, SMOTE è stato applicato tra la classe Target e la classe Non-Target, raggiungendo un totale di 146.896 samples per entrambe, migliorando così la rappresentazione della classe minoritaria.

Per quanto riguarda la classificazione multiclasse, SMOTE è stato utilizzato per bilanciare tutte le sottoclassi della classe Target, portandole a 9.181 samples ciascuna. A questo punto, la classe Target è composta da un set di addestramento con 16 sottoclassi (tutte bilanciate), uno di validazione con 6 sottoclassi e uno di test con 7 sottoclassi. Durante l'applicazione di SMOTE, il parametro  $k\_neighbors$  è stato testato con diversi valori per determinare l'adeguato numero di vicini da considerare nella generazione dei samples sintetici. È stato scelto un valore relativamente basso di  $k\_neighbors$ , poiché molte delle sottoclassi erano poco rappresentate nel dataset. Inoltre, le classi con un solo sample sono state rimosse, in quanto non consentivano il corretto funzionamento di SMOTE. Questo approccio ha preservato meglio le caratteristiche di ciascuna sottoclasse minoritaria, evitando la creazione di samples sintetici troppo distanti dai dati reali.

Per migliorare ulteriormente la rappresentazione delle classi minoritarie e ridurre il rischio di overfitting, SMOTE è stato applicato in entrambe le classificazioni. Questo ha contribuito a mantenere la diversità dei segnali bioacustici e ha consentito al modello di apprendere caratteristiche distintive per ciascuna categoria (3) (Chawla et al., 2002). I risultati ottenuti indicano un aumento significativo nella precisione e nella recall per le classi minoritarie, contribuendo a una performance complessiva più robusta.

## 9. Integrazione delle Features

La combinazione delle features descritte in precedenza si è dimostrata efficace nel migliorare l'accuratezza della classificazione dei segnali bioacustici rispetto all'uso di singole features. Studi recenti hanno evidenziato che la selezione e l'ottimizzazione di features come le MFCC (Mel-frequency cepstral coefficients), combinate con centroidi spettrali e skewness, possono incrementare significativamente le performance di modelli di machine learning come Support Vector Machines e Random Forests (Malfante et al., 2018). L'integrazione di queste caratteristiche consente di costruire un modello robusto, in grado di distinguere non solo tra suoni antropogenici e bioacustici, ma anche di classificare con precisione le specie marine coinvolte. Questo approccio fornisce un

sistema di monitoraggio acustico avanzato e altamente affidabile.

Sono state inoltre utilizzate numerose features numeriche tratte da BirdNet, tra cui gli MFCC, lo Zero Crossing Rate, il Centroid Spettrale e i Chroma. Queste caratteristiche sono state selezionate per la loro capacità di catturare importanti aspetti acustici, risultando efficaci nella classificazione dei segnali.

Successivamente, è stato considerato anche un altro set di features proposto da Dhamodaran et al. nello studio "A Survey on Audio Feature Extraction for Automatic Music Genre Classification", ampliando ulteriormente la gamma di caratteristiche analizzate per migliorare le prestazioni del modello.

## 10. Addestramento

Per l'addestramento del modello, sono stati utilizzati tre algoritmi di machine learning: Random Forest, Support Vector Machine (SVM) e LightGBM. Ognuno di questi modelli è stato scelto per le proprie caratteristiche distintive e la capacità di affrontare problemi di classificazione in contesti complessi come quello dei segnali bioacustici.

- **Random Forest:** Questo modello è un ensemble di alberi decisionali che utilizza il metodo del bagging per migliorare la precisione e ridurre il rischio di overfitting. La Random Forest è particolarmente efficace nel gestire dataset con molte variabili, consentendo di catturare relazioni non lineari nei dati. Nel caso specifico, il modello è stato implementato con 100 alberi decisionali, impostando un valore di *random seed* pari a 42 per garantire la riproducibilità dei risultati. L'addestramento è stato eseguito iterativamente grazie al parametro *warm\_start=True*, il quale ha permesso di aggiungere progressivamente un albero per ogni iterazione, mantenendo i modelli precedentemente addestrati. Questa strategia ha consentito di ottimizzare il modello incrementando gradualmente il numero di alberi da 1 a 100.
- **Support Vector Machine (SVM):** Le SVM (Support Vector Machine) sono un potente strumento di classificazione, progettato per trovare l'iperpiano ottimale che separa le classi nel piano multidimensionale. Sono particolarmente utili in scenari ad alta dimensione, dove possono essere utilizzate diverse funzioni di kernel per gestire la non linearità. L'SVM è stato configurato per utilizzare la loss function '*hinge*', che è particolarmente efficace per problemi di classificazione binaria. Il parametro *random\_state=42* è impostato per garantire la riproducibilità dei risultati.
- **LightGBM:** Questo algoritmo di boosting è progettato per essere efficiente in termini di tempo e spazio. Utilizza un approccio di gradient boosting basato su alberi, ed è in grado di gestire grandi

dataset e di ottenere prestazioni elevate grazie alla sua capacità di ridurre il numero di dati necessari per l'addestramento. Il modello di LightGBM è inizializzato con il parametro *random\_state* impostato a un valore fisso (42) per garantire la riproducibilità dei risultati, assicurando che ogni esecuzione dell'addestramento produca gli stessi risultati, a prescindere dalla variabilità intrinseca dei dati o delle condizioni di esecuzione.

L'approccio di addestramento utilizzato prevede la suddivisione del set di dati in blocchi incrementali, determinati dall'argomento *n\_steps=10*, che specifica il numero di passaggi per l'addestramento. Questo metodo consente al modello di apprendere progressivamente, adattandosi alle informazioni contenute in ciascun blocco di dati in modo sistematico. Durante ogni iterazione, il modello viene addestrato utilizzando un sottoinsieme crescente di dati, che contribuisce a migliorare l'efficienza dell'addestramento e a ottimizzare le prestazioni, specialmente in scenari con grandi volumi di dati.

Questi modelli sono stati addestrati utilizzando il set di dati bilanciato attraverso SMOTE, permettendo una valutazione equa delle loro prestazioni nella classificazione dei segnali bioacustici.

## 11. Risultati

In questa sezione vengono riportati i risultati ottenuti dai tre modelli di classificazione implementati. L'obiettivo di questa analisi è valutare l'efficacia di ciascun modello nel classificare correttamente il dataset, utilizzando tre diversi insiemi di feature selezionate dalla letteratura. Questi insiemi di feature sono stati costruiti con l'intento di catturare diversi aspetti rilevanti dei dati, al fine di confrontare le prestazioni dei modelli su diverse rappresentazioni del problema. Le metriche di performance riportate considerano sia i risultati ottenuti durante la fase di validazione che quelli osservati sul set di test, permettendo di confrontare il comportamento dei modelli su dati mai visti prima. Questo garantisce che le performance siano robuste e generalizzabili. Per misurare le prestazioni di ciascun modello, sono state utilizzate le seguenti metriche:

-**Precision:** indica la proporzione di predizioni corrette tra quelle assegnate alla classe positiva.

-**Recall:** misura la capacità del modello di identificare correttamente tutti i samples della classe positiva.

-**F1-score:** una combinazione bilanciata di precision e recall, utile per valutare le prestazioni quando c'è uno squilibrio tra le classi.

-**Accuracy:** rappresenta la percentuale complessiva di predizioni corrette rispetto al totale dei samples.

Inoltre, sono state considerate sia la **macro media** (la media semplice delle metriche per ciascuna classe, indipendentemente dalla distribuzione delle classi), sia la **media ponderata** (che considera la proporzione delle



classi). Questo approccio consente di valutare come i modelli gestiscono eventuali squilibri tra le classi. Nel processo di valutazione, sono stati utilizzati sia un set di validazione che un set di test per stimare le capacità di generalizzazione dei modelli.

**Set di validazione:** utilizzato durante il processo di addestramento per ottimizzare i modelli e selezionare i parametri migliori. Le metriche calcolate su questo set aiutano a monitorare l'adattamento dei modelli ai dati e a evitare fenomeni di overfitting.

**Set di test:** utilizzato esclusivamente dopo l'addestramento e la selezione dei parametri. Serve per valutare l'efficacia finale del modello su dati completamente nuovi, fornendo una stima realistica delle sue prestazioni. I risultati ottenuti dai modelli su ciascun set di feature sono riportati nelle tabelle che seguono, con le metriche calcolate su entrambi i set di validazione e di test. Le differenze tra le due valutazioni possono indicare la capacità del modello di generalizzare correttamente oppure segnalare fenomeni di overfitting o underfitting.

Le features individuate sono state suddivise in 3 set di dati:

**Table 1**  
Set di Dati Comparativi

Primo Set (11)	
Spectral Centroid Mean	Spectral Bandwidth RMS
Standard Deviation	Skewness
Kurtosis	Shannon Entropy
Renyi Entropy	Rate of Attack
Rate of Decay	Threshold Crossings
Silence Ratio	Mean
Max Over Mean	Min Over Mean
Energy-Measurements	
Secondo Set (BirdNet)	
MFCC 1	MFCC 2
MFCC 3	MFCC 4
MFCC 5	MFCC 6
MFCC 7	MFCC 8
MFCC 9	MFCC 10
MFCC 11	MFCC 12
MFCC 13	ZCR
Spectral Centroid	Spectral Bandwidth
Chroma 1	Chroma 2
Chroma 3	Chroma 4
Chroma 5	Chroma 6
Chroma 7	Chroma 8
Chroma 9	Chroma 10
Chroma 11	Chroma 12
Terzo Set (19)	
MFCC 1	MFCC 2
MFCC 3	MFCC 4
MFCC 5	MFCC 6
MFCC 7	MFCC 8
MFCC 9	MFCC 10
MFCC 11	MFCC 12
MFCC 13	ZCR
Spectral Contrast	Tonnetz 1
Tonnetz 2	Tonnetz 3
Tonnetz 4	Tonnetz 5
Tonnetz 6	Chroma 1
Chroma 2	Chroma 3
Chroma 4	Chroma 5
Chroma 6	Chroma 7
Chroma 8	Chroma 9
Chroma 10	Chroma 11
Chroma 12	Tempo

	Precision	Recall	F1-score	Support
macro avg(val)	0.99	0.99	0.99	4333
weighted avg(val)	0.99	0.99	0.99	4333
macro avg(test)	0.94	0.79	0.84	6234
weighted avg(test)	0.93	0.92	0.92	6234

*Accuratezza sul set di validazione: 0.9914*

*Accuratezza sul set di test: 0.9231*

**Table 2**

Risultati addestramento binario con Random Forest (primo set.)

La Random Forest ha ottenuto prestazioni eccellenti durante la validazione, con precisione, richiamo e F1-score prossimi a 1, indicando un'elevata accuratezza nel classificare le istanze positive e negative. L'accuratezza sul set di validazione è pari a 99.14%, confermando l'efficacia del modello. Sul set di test, il modello ha ottenuto un'accuratezza del 92.3%, mostrando un leggero calo nelle performance, soprattutto nel richiamo (0.79 per la media macro). Tuttavia, i valori di precisione e F1-score rimangono elevati, suggerendo che il modello mantiene buone prestazioni, nonostante il set di test presenti alcune sfide aggiuntive.

	Precision	Recall	F1-score	Support
macro avg(val)	0.81	0.87	0.83	4333
weighted avg(val)	0.89	0.87	0.88	4333
macro avg(test)	0.82	0.79	0.80	6234
weighted avg(test)	0.88	0.89	0.88	6234

*Accuratezza sul set di validazione: 0.8786*

*Accuratezza sul set di test: 0.8928*

**Table 3**

Risultati addestramento binario con SVM (primo set).

L'SVM ha ottenuto buone prestazioni sul set di validazione, con un'accuratezza di 87.86%. Tuttavia, il richiamo nella media macro (0.87) è inferiore rispetto alla Random Forest, il che indica che il modello potrebbe avere difficoltà nel catturare alcune classi minoritarie. Sul set di test, l'accuratezza è leggermente migliorata (89.28%), con prestazioni equilibrate tra precisione e richiamo. Questo suggerisce che l'SVM è in grado di generalizzare bene su dati non visti, ma potrebbe ancora avere margini di miglioramento nel trattare la variabilità del dataset.

	Precision	Recall	F1-score	Support
macro avg(val)	0.98	0.99	0.98	4333
weighted avg(val)	0.99	0.99	0.99	4333
macro avg(test)	0.98	0.98	0.98	6234
weighted avg(test)	0.99	0.99	0.99	6234

*Accuratezza sul set di validazione: 0.9892*

*Accuratezza sul set di validazione: 0.9885*

**Table 4**

Risultati addestramento binario con LightGBM (primo set).

Il modello LightGBM ha ottenuto risultati eccellenti, con un F1-score molto alto (0.98-0.99), segno di un ottimo bilanciamento tra precisione e richiamo. L'accuratezza sul set di validazione è del 98.92%, confermando la capacità del modello di gestire efficacemente le diverse classi. I risultati sul set di test, con un'accuratezza di 98.85%, confermano che il modello generalizza molto bene, mantenendo prestazioni quasi identiche a quelle ottenute durante la validazione, dimostrando un'elevata robustezza.

	Precision	Recall	F1-score	Support
macro avg	0.98	0.99	0.99	4333
weighted avg	0.99	0.99	0.99	4333
macro avg(test)	0.98	0.96	0.97	6234
weighted avg(test)	0.98	0.98	0.98	6234

*Accuratezza sul set di validazione: 0.9903*

*Accuratezza sul set di test: 0.9807*

**Table 5**

Risultati addestramento binario con Random Forest (secondo set).

Le performance del modello Random Forest sul secondo set sono elevate, con valori di F1-score e precisione prossimi a 1 sia per la media macro che pesata. L'accuratezza sul set di validazione è del 99.03%, indicando un'eccellente capacità di classificare correttamente le istanze. Sul set di test, l'accuratezza è leggermente inferiore (98.07%), ma le metriche restano coerenti con quelle del set di validazione, suggerendo che il modello è altamente affidabile e stabile.

	Precision	Recall	F1-score	Support
macro avg(val)	0.84	0.84	0.84	4333
weighted avg(val)	0.89	0.89	0.89	4333
macro avg(test)	0.78	0.73	0.75	6234
weighted avg(test)	0.85	0.86	0.86	6234

*Accuratezza sul set di validazione: 0.8975*

*Accuratezza sul set di test: 0.8681*

**Table 6**

Risultati addestramento binario con SVM (secondo set).

Il modello SVM ha mostrato buone prestazioni complessive con un'accuratezza del 89.75% sul set di validazione. Tuttavia, il richiamo nella media macro (0.84) indica che alcune classi potrebbero non essere state catturate in modo ottimale. I risultati sul set di test (accuratezza 86.81%) confermano che il modello ha prestazioni leggermente peggiori su dati non visti, suggerendo che potrebbe essere più sensibile alla variabilità del dataset rispetto ad altri modelli.

	Precision	Recall	F1-score	Support
macro avg(val)	0.98	0.97	0.98	4333
weighted avg(val)	0.98	0.98	0.98	4333
macro avg(test)	0.98	0.97	0.97	6234
weighted avg(test)	0.99	0.99	0.99	6234

*Accuratezza sul set di validazione: 0.9836*

*Accuratezza sul set di test: 0.9854*

**Table 7**

Risultati addestramento binario con Light GBM (secondo set). Il LightGBM ha ottenuto ottimi risultati con un'accuratezza del 98.36% sul set di validazione e metriche elevate per precisione, richiamo e F1-score. Questi risultati indicano che il modello ha un'elevata capacità di generalizzare, mantenendo un'elevata accuratezza (98.54%) anche sul set di test. Le prestazioni coerenti tra i due set confermano la capacità del modello di adattarsi bene ai dati, senza overfitting.

	Precision	Recall	F1-score	Support
macro avg(val)	0.98	0.99	0.98	4333
weighted avg(val)	0.99	0.99	0.99	4333
macro avg(test)	0.97	0.94	0.96	6234
weighted avg(test)	0.97	0.97	0.97	6234

*Accuratezza sul set di validazione: 0.9898*

*Accuratezza sul set di test: 0.9748*

**Table 8**

Risultati addestramento binario con Random Forest (terzo set). La Random Forest ha dimostrato un'eccellente capacità di classificare le istanze del terzo set, con precisione, richiamo e F1-score molto elevati. L'accuratezza sul set di validazione è del 98.98%, confermando la robustezza del modello. Sul set di test, il modello ha ottenuto un'accuratezza leggermente inferiore (97.48%), ma con metriche simili, dimostrando una buona generalizzazione anche su dati non visti.

	Precision	Recall	F1-score	Support
macro avg(val)	0.82	0.80	0.81	4333
weighted avg(val)	0.87	0.87	0.87	4333
macro avg(test)	0.76	0.73	0.74	6234
weighted avg(test)	0.85	0.85	0.85	6234

*Accuratezza sul set di validazione: 0.8792*

*Accuratezza sul set di test: 0.8591*

**Table 9**

Risultati addestramento binario con SVM (terzo set). Il modello SVM ha mostrato una performance inferiore rispetto agli altri modelli, con un'accuratezza del 87.92% sul set di validazione. Le metriche di richiamo e F1-score (0.73 e 0.74) sul set di test indicano una leggera riduzione delle prestazioni rispetto alla validazione, con un'accuratezza del 85.91%. Il modello potrebbe avere più difficoltà nel trattare la variabilità del dataset rispetto a soluzioni come la Random Forest o il LightGBM.

	Precision	Recall	F1-score	Support
macro avg(val)	0.98	0.99	0.99	4333
weighted avg(val)	0.99	0.99	0.99	4333
macro avg(test)	0.98	0.96	0.97	6234
weighted avg(test)	0.98	0.98	0.98	6234

*Accuratezza sul set di validazione: 0.9915*

*Accuratezza sul set di test: 0.9809*

**Table 10**

Risultati addestramento binario con Light GBM (terzo set). I risultati ottenuti indicano che il modello LightGBM ha dimostrato un'eccellente capacità di classificare correttamente le istanze del terzo set di dati. La sua alta precisione, recall e F1-score suggeriscono che il modello è in grado di distinguere con grande affidabilità tra le due classi. I risultati sul set di test risultano i medesimi rispetto a quelli sul set di validazione.

	Precision	Recall	F1-score	Support
macro avg(val)	0.29	0.22	0.24	3427
weighted avg(val)	0.37	0.36	0.36	3427
macro avg(test)	0.39	0.43	0.41	5119
weighted avg(test)	0.42	0.39	0.40	5119

*Accuratezza sul set di validazione: 0.3580*

*Accuratezza sul set di test: 0.3970*

**Table 11**

Risultati addestramento multiclasse con Random Forest (primo set).

Il modello Random Forest per la classificazione multiclasse non ha ottenuto risultati soddisfacenti, con metriche basse per precisione, richiamo e F1-score. L'accuratezza del 35.80% sul set di validazione suggerisce una scarsa capacità di distinguere tra le classi. Anche sul set di test, il modello ha riportato risultati simili (accuratezza 39.70%), confermando la difficoltà nel generalizzare in contesti multiclasse complessi.

	Precision	Recall	F1-score	Support
macro avg(val)	0.10	0.09	0.04	3427
weighted avg(val)	0.28	0.18	0.08	3427
macro avg(test)	0.10	0.15	0.11	5119
weighted avg(test)	0.07	0.20	0.08	5119

*Accuratezza sul set di validazione: 0.1806*

*Accuratezza sul set di test: 0.2026*

**Table 12**

Risultati addestramento multiclasse con SVM (primo set). Le prestazioni dell'SVM sono state particolarmente deludenti, con valori molto bassi per tutte le metriche. L'accuratezza sul set di validazione è stata di appena 18.06%, e sul set di test è migliorata leggermente a 20.26%. Questo indica che il modello fatica a identificare correttamente le diverse classi, con performance che restano molto limitate su entrambi i set.

	Precision	Recall	F1-score	Support
macro avg(val)	0.38	0.61	0.28	3427
weighted avg(val)	0.36	0.36	0.36	3427
macro avg(test)	0.31	0.68	0.31	5119
weighted avg(test)	0.41	0.39	0.39	5119

*Accuratezza sul set di validazione: 0.3645*

*Accuratezza sul set di test: 0.3862*

**Table 13**

Risultati addestramento multiclasse con Light GBM (primo set).

Il modello LightGBM ha mostrato un leggero miglioramento rispetto ai modelli precedenti, con un'accuratezza del 36.45% sul set di validazione. Le metriche di richiamo sono particolarmente elevate rispetto a precisione e F1-score, suggerendo una maggiore capacità di catturare le classi rispetto ai modelli come Random Forest e SVM. I risultati sul set di test (accuratezza 38.62%) sono in linea con quelli della validazione, ma rimangono comunque insufficienti per considerare il modello pienamente affidabile.

	Precision	Recall	F1-score	Support
macro avg(val)	0.50	0.54	0.52	3427
weighted avg(val)	0.53	0.54	0.53	3427
macro avg(test)	0.60	0.56	0.57	5119
weighted avg(test)	0.56	0.55	0.55	5119

*Accuratezza sul set di validazione: 0.5404*

*Accuratezza sul set di test: 0.5458*

**Table 14**

Risultati addestramento multiclasse con Random Forest (secondo set).

Il modello Random Forest ha mostrato miglioramenti significativi rispetto agli altri modelli multiclasse, con un'accuratezza del 54.04% sul set di validazione. Le metriche di precisione, richiamo e F1-score indicano che il modello è in grado di distinguere meglio le classi rispetto alle versioni precedenti. Anche sul set di test, l'accuratezza è simile (54.58%), dimostrando una capacità di generalizzare leggermente migliore rispetto agli altri modelli.

	Precision	Recall	F1-score	Support
macro avg(val)	0.26	0.21	0.19	3427
weighted avg(val)	0.43	0.33	0.29	3427
macro avg(test)	0.21	0.25	0.19	5119
weighted avg(test)	0.36	0.24	0.20	5119

*Accuratezza sul set di validazione: 0.3309*

*Accuratezza sul set di test: 0.2352*

**Table 15**

Risultati addestramento multiclasse con SVM (secondo set).

I risultati ottenuti con l'addestramento del modello SVM sul secondo set di dati multiclasse mostrano una performance modesta. La precisione, il richiamo e l'F1-score medi sono bassi, con un F1-score ponderato di 0.08 sul set di validazione, e simili valori sul set di test. Questo suggerisce che il modello ha grosse difficoltà a distinguere correttamente le diverse classi. L'accuratezza complessiva sul set di validazione è pari a 0.18, con una leggera crescita sul set di test (0.20). Questi risultati confermano che il modello SVM non è adatto a questo specifico task multiclasse, presentando un'accuratezza molto bassa sia in validazione che test.

	Precision	Recall	F1-score	Support
macro avg(val)	0.60	0.60	0.49	3427
weighted avg(val)	0.56	0.56	0.55	3427
macro avg(test)	0.63	0.63	0.63	5119
weighted avg(test)	0.60	0.58	0.59	5119

*Accuratezza sul set di validazione: 0.5579*

*Accuratezza sul set di test: 0.5814*

**Table 16**

Risultati addestramento multiclasse con Light GBM(secondo set).

Il modello Light GBM applicato al secondo set di dati multiclasse ha ottenuto prestazioni migliori rispetto ai modelli precedenti. I valori di F1-score, precisione e recall medi suggeriscono un equilibrio tra le classi, con un F1-score ponderato di 0.39. Tuttavia, la precisione e il recall individuali indicano che il modello potrebbe avere difficoltà con alcune classi specifiche. L'accuratezza complessiva sul set di validazione è di 0.3645, mentre sul set di test il valore è molto simile, 0.3862. Sebbene questi risultati siano ancora distanti da una performance ottimale, il modello sembra essere leggermente più robusto rispetto a SVM e Random Forest nel contesto multiclasse, confermando la sua capacità di gestire dataset complessi.

	Precision	Recall	F1-score	Support
macro avg(val)	0.35	0.30	0.31	3427
weighted avg(val)	0.52	0.52	0.51	3427
macro avg(test)	0.50	0.42	0.42	5119
weighted avg(test)	0.57	0.56	0.56	5119

*Accuratezza sul set di validazione: 0.5182*

*Accuratezza sul set di test: 0.5589*

**Table 17**

Risultati addestramento multiclasse con Random Forest(terzo set).

I risultati del modello Random Forest per il terzo set di dati multiclasse indicano un significativo miglioramento delle prestazioni rispetto ai modelli precedenti. La precisione, il richiamo e l'F1-score medi sono notevolmente più elevati, con un F1-score ponderato di 0.57. Questi risultati suggeriscono che il modello ha acquisito una maggiore capacità di distinguere tra le diverse classi rispetto agli altri set. L'accuratezza sul set di validazione è 0.5404, con un leggero miglioramento sul set di test (0.5458). Ciò dimostra che il modello è in grado di generalizzare meglio, anche se potrebbero ancora esserci margini di miglioramento per una classificazione più accurata su tutte le classi.

	Precision	Recall	F1-score	Support
macro avg(val)	0.16	0.15	0.14	3427
weighted avg(val)	0.35	0.35	0.33	3427
macro avg(test)	0.32	0.38	0.34	5119
weighted avg(test)	0.48	0.46	0.46	5119

*Accuratezza sul set di validazione: 0.3510*

*Accuratezza sul set di test: 0.4604*

**Table 18**

Risultati addestramento multiclasse con SVM(terzo set).  
I risultati ottenuti con l'SVM sul terzo set di dati sono piuttosto deludenti. L'accuratezza complessiva è solo del 35,10%, indicando che il modello ha classificato correttamente meno di un terzo delle istanze. Risultati migliori per il set di test.

	Precision	Recall	F1-score	Support
macro avg(val)	0.59	0.67	0.55	3427
weighted avg(val)	0.53	0.53	0.52	3427
macro avg(test)	0.55	0.68	0.54	5119
weighted avg(test)	0.61	0.59	0.60	5119

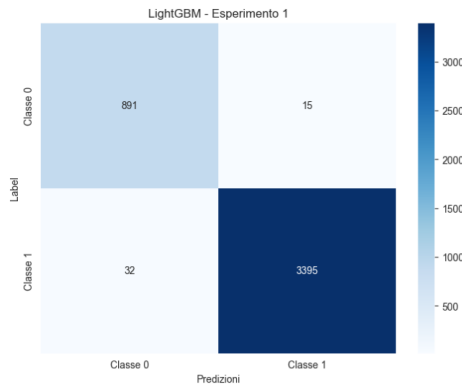
*Accuratezza sul set di validazione: 0.5279*

*Accuratezza sul set di test: 0.5945*

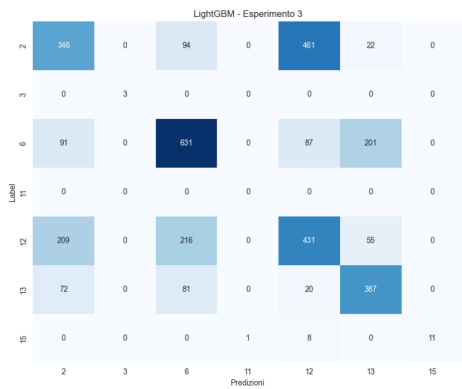
**Table 19**

Risultati addestramento multiclasse con Light GBM(terzo set).  
La tabella mostra che il modello Light GBM, applicato al terzo set di dati multiclasse, ha ottenuto i migliori risultati finora per quanto riguarda questo task. La precisione, il recall e l'F1-score medi sono relativamente alti, con un F1-score ponderato di 0.55 e una buona capacità di bilanciamento tra le classi. L'accuratezza sul set di validazione è 0.5450, mentre sul set di test risulta leggermente più bassa, 0.5407. Questo suggerisce che il modello è in grado di generalizzare bene a nuovi dati, sebbene ci siano ancora alcune difficoltà nell'identificare correttamente le istanze appartenenti a tutte le classi.





**Figure 6:** Matrice di confusione primo set classificazione binaria con LightGBM



**Figure 7:** Matrice di confusione terzo set classificazione multi-classe con LightGBM

## 12. Esperimenti sui modelli migliori

Dopo aver analizzato le prestazioni dei modelli, sono stati identificati i due migliori: LightGBM per la classificazione binaria (primo set), con un valore di accuracy sul set di test pari a 0.9885, e LightGBM per la classificazione multiclasse (terzo set), con un valore di accuracy di 0.5945. Per ottimizzare ulteriormente le prestazioni, sono stati condotti diversi esperimenti modificando il sample rate e fornendo nuovi set di dati ai modelli. Di seguito gli esperimenti riguardanti la classificazione binaria. Il primo esperimento, effettuato su LightGBM utilizzando il primo set di dati con un sample rate di 44100 Hz, ha prodotto i seguenti risultati:

	Precision	Recall	F1-score	Support
macro avg(val)	0.96	0.95	0.95	4333
weighted avg(val)	0.98	0.98	0.98	4333
macro avg(test)	0.94	0.82	0.86	6234
weighted avg(test)	0.93	0.93	0.92	6234

*Accuratezza sul set di validazione: 0.9811*

*Accuratezza sul set di test: 0.9301*

Successivamente, è stato effettuato un secondo esperimento sempre per la classificazione binaria con LightGBM, utilizzando il primo set di dati ma con un sample rate di 192000 Hz. I risultati ottenuti sono i seguenti:

	Precision	Recall	F1-score	Support
macro avg(val)	0.97	0.97	0.97	4333
weighted avg(val)	0.98	0.98	0.98	4333
macro avg(test)	0.92	0.78	0.83	6234
weighted avg(test)	0.92	0.91	0.91	6234

*Accuratezza sul set di validazione: 0.9785*

*Accuratezza sul set di test: 0.9143*

Infine, è stato condotto un terzo esperimento per la classificazione binaria con LightGBM sul primo set di dati, questa volta utilizzando un sample rate di 384000 Hz. I risultati sono stati i seguenti:

	Precision	Recall	F1-score	Support
macro avg(val)	0.97	0.97	0.97	4333
weighted avg(val)	0.98	0.98	0.98	4333
macro avg(test)	0.93	0.79	0.84	6234
weighted avg(test)	0.92	0.92	0.91	6234

*Accuratezza sul set di validazione: 0.9815*

*Accuratezza sul set di test: 0.9211*

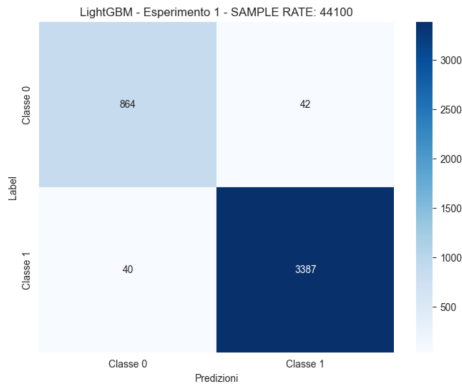
Di seguito i risultati degli esperimenti della classificazione multiclasse utilizzando il terzo set di features su LightGBM.

Esperimento con sample rate a 44100 Hz:

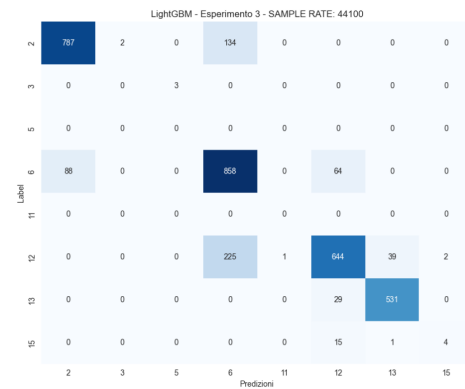
	Precision	Recall	F1-score	Support
macro avg(val)	0.51	0.69	0.46	3427
weighted avg(val)	0.83	0.82	0.82	3427
macro avg(test)	0.52	0.81	0.43	5119
weighted avg(test)	0.78	0.77	0.77	5119

*Accuratezza sul set di validazione: 0.8240*

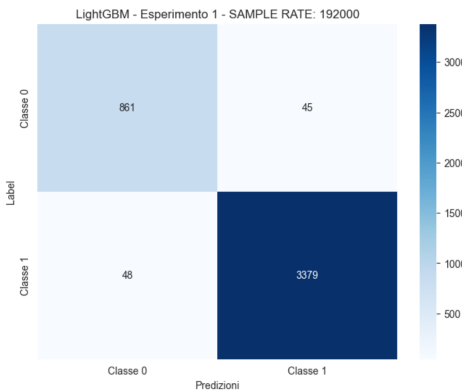
*Accuratezza sul set di test: 0.7675*



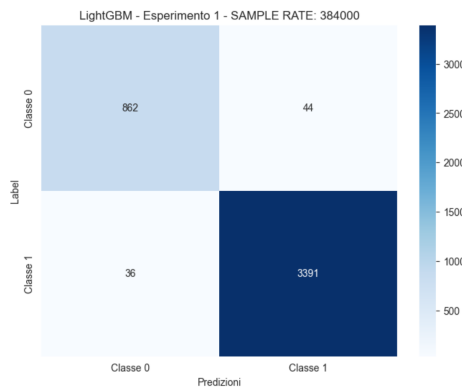
**Figure 8:** Matrice di confusione esperimento 1 (features di Malfante) classificazione binaria 44100Hz con LightGBM



**Figure 11:** Matrice di confusione esperimento 3 (features di Dhamodaran) classificazione multiclasse 44100Hz con LightGBM



**Figure 9:** Matrice di confusione esperimento 1 (features di Malfante) classificazione binaria 192000Hz con LightGBM



**Figure 10:** Matrice di confusione esperimento 1 (features di Malfante) classificazione binaria 384000Hz con LightGBM

### 13. Conclusioni

In conclusione, l'analisi dei risultati ottenuti evidenzia che per il task di classificazione binaria, i modelli LightGBM e Random Forest si sono dimostrati i più performanti, raggiungendo accuratèzze superiori al 98% e mostrando un ottimo bilanciamento tra precision, recall e F1-score. Entrambi i modelli si sono rivelati adatti per applicazioni in cui è fondamentale garantire elevate performance di classificazione. Al contrario, l'SVM, pur avendo fornito

risultati discreti, è generalmente meno performante rispetto ai due modelli precedenti. Per quanto riguarda la classificazione multiclasse, il LightGBM si è confermato il modello più efficace, con accuratèzze superiori rispetto a Random Forest e SVM, dimostrandosi più adatto per gestire la complessità di questo tipo di task. In sintesi, mentre LightGBM e Random Forest sono altamente consigliati per task di classificazione binaria, il LightGBM si distingue come la scelta più robusta anche per la classificazione multiclasse.

I risultati degli esperimenti di classificazione binaria con LightGBM, effettuati variando il sample rate del dataset, dimostrano che il resampling a frequenze diverse da 96.000 Hz non migliora significativamente le prestazioni del modello. In particolare, ridurre il sample rate a 44.100 Hz ha causato una diminuzione dell'accuratèzza sul set di test, scesa da 0,9885 a 0,9301. Un comportamento simile è stato osservato anche con l'aumento del sample rate a 192.000 Hz e 384.000 Hz, con un'accuratèzza ulteriormente ridotta a 0,9143 e 0,9211 rispettivamente.

Nella classificazione multiclasse, l'uso di un sample rate di 44.100 Hz ha portato a miglioramenti significativi nelle prestazioni del modello. L'accuratèzza sul set di test è aumentata da 0.5945 a 0.7675, mentre sul set di validazione è passata da 0.5279 a 0.8240.

I risultati ottenuti evidenziano l'importanza della scelta della frequenza di campionamento nell'ottimizzazione delle prestazioni.

### 14. Sviluppi futuri

Gli sviluppi futuri di questo progetto potrebbero concentrarsi su diversi aspetti chiave per migliorare l'accuratèzza e l'efficacia del modello di classificazione. In primo luogo, sarebbe opportuno utilizzare un dataset più ampio e diversificato, che includa una maggiore varietà di suoni bioacustici e antropogenici, raccolti in diverse condizioni ambientali. Un dataset più rappresentativo migliorerebbe la capacità del modello di generalizzare e adattarsi a una più ampia gamma di scenari reali. Inoltre,

l'integrazione di combinazioni avanzate di caratteristiche numeriche (features) acustiche potrebbe permettere una rappresentazione più dettagliata dei segnali audio, consentendo di catturare meglio le peculiarità dei suoni bioacustici e antropogenici. L'uso di tecniche di feature selection e extraction, come l'analisi delle componenti principali (PCA) o metodi basati su deep learning, potrebbe ottimizzare la rappresentazione dei segnali, migliorando l'accuratezza del modello. Infine, l'impiego di modelli di machine learning più avanzati e complessi, come reti neurali convoluzionali (CNN) o modelli basati su apprendimento profondo, potrebbe incrementare le prestazioni del sistema. Un approccio ensemble, che combini diversi algoritmi di classificazione, potrebbe ulteriormente raffinare i risultati, sfruttando la complementarità tra i vari modelli per ottenere una maggiore robustezza nella distinzione tra suoni antropogenici e bioacustici. Questi sviluppi futuri offriranno un potenziale significativo per migliorare la capacità del sistema di supportare la gestione sostenibile degli ecosistemi marini attraverso una più accurata identificazione dei segnali acustici.

## 15. Bibliography

### References

- [1] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- [2] Blumstein, D. T., Mennill, D. J., Clemins, P., Girod, L., Yao, K., Patricelli, G., ... & Kirschel, A. N. (2011). Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *Journal of Applied Ecology*, 48(3), 758-767.
- [3] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321-357.
- [4] Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory*. John Wiley & Sons.
- [5] Figueroa, L., Belloni, M., & Abascal-Mena, R. (2015). Statistical modeling of cetacean click properties for automatic classification. *Journal of the Acoustical Society of America*, 137(3), 1025-1034.
- [6] Gibb, R., Browning, E., Glover-Kapfer, P., & Jones, K. E. (2019). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution*, 10(2), 169-185.
- [7] Giannakopoulos, T., & Piskakis, A. (2014). *Introduction to Audio Analysis: A MATLAB® Approach*. Academic Press.
- [8] Hildebrand, J. A. (2009). Anthropogenic and natural sources of ambient noise in the ocean. *Marine Ecology Progress Series*, 395, 5-20.
- [9] Kahl, S., Wood, C. M., Eibl, M., & Klinck, H. (2021). BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, 61, 101236.
- [10] Lerch, A. (2012). *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. John Wiley & Sons.
- [11] Malfante, M., Mars, J. I., Dalla Mura, M., & Gervaise, C. (2018). Automatic fish sounds classification in real-life marine environments: Performance evaluation and perspectives. *Journal of the Acoustical Society of America*, 143(5), 2834-2845.
- [12] Mellinger, D. K., & Clark, C. W. (2000). Recognizing transient low-frequency whale sounds by spectrogram correlation. *The Journal of the Acoustical Society of America*, 107(6), 3518-3529.
- [13] Mermelstein, P. (1976). Distance measures for speech recognition, psychological and instrumental. *Pattern Recognition and Artificial Intelligence*, 116, 374-388.
- [14] Popescu, M., Ehrlich, R., & Xu, W. (2009). Detection and classification of acoustic events for in-home monitoring of an elder person living alone. In *Proceedings of the 6th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* (pp. 314-319).
- [15] Salamon, J., Bello, J. P., Farnsworth, A., & Rosenheim, K. (2017). Feature learning with convolutional neural networks for bioacoustics. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2662-2666). IEEE.
- [16] Sharma, R., Agarwal, S., & Jain, R. (2020). Acoustic signal classification using hybrid features and machine learning techniques. *Journal of Ambient Intelligence and Humanized Computing*, 11(3), 1171-1180.
- [17] Sueur, J., Aubin, T., & Simonis, C. (2008). Equipment review: Seewave, a free modular tool for sound analysis and synthesis. *Bioacoustics*, 18(2), 213-226.
- [18] Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5), 293-302.
- [19] Dhamodaran, P., Balakrishnan, V., & Dharmavaram, A. (2023). A Survey on Audio Feature Extraction for Automatic Music Genre Classification. *International Journal of Recent Technology and Engineering*, 12(3), 45-52.