

Gépi tanuláson (ML AI) alapuló adatosztályozás a víz alatti hangok és zajok elemzésére

Machine Learning (ML AI) based data classification for underwater sound and noise analysis

ARADI Attila¹, Dr. VARGA Attila Károly²

¹ ION-technik Kft. Aradi.Atila@ion-technik.hu

² Miskolci Egyetem, Attila.Varga@uni-miskolc.hu

Abstract

In this article, we introduce a machine learning method rooted in artificial intelligence (AI) and neural network (NN) frameworks for categorizing underwater acoustics, marine ambient noises, and marine mammal vocalizations (such as those of dolphins and other cetaceans) captured via hydrophones. The gathering of such audio data yields an extensive collection of underwater sound samples. Manually analyzing these recordings can be tedious, prone to mistakes, and may lead to oversight, especially given the repetitive nature of the task. Bearing these factors in mind, we explore the potential of employing AI for refining and categorizing these audio samples, aiming to identify specific marine species like dolphins and cetaceans, as well as pinpointing the type of vessel for ship-generated noises.

Kivonat

Ebben a tanulmányban egy mesterséges intelligencián (AI) azaz neurális hálózaton (NN) alapuló gépi tanulási technikát mutatunk be, és annak alkalmazását a víz alatti hangok, víz alatti zajok és tengeri emlősök (delfinek, cetfélék) hidrofonokkal rögzített víz alatti hangjainak osztályozására. Az adatgyűjtési folyamat során nagy mennyiségű víz alatti hangfelvétel keletkezik, amelyek emberi feldolgozására igen időigényes, és a munka monoton jellege miatt nagy az emberi hiba és az emberi figyelmetlenség lehetősége. Ezeket a szempontokat figyelembe véve felmerült a mesterséges intelligencia alkalmazásának lehetősége a hangfelvételek utólagos feldolgozására, a hangok osztályozására, a cetfélék és delfinek esetében a fajfelismerés szintjéig, a hajóforgalom zaja esetében pedig a hajótípus-felismerés szintjéig.

Kulcsszavak: hidrofon, bioakusztika, gépi tanulás, mesterséges intelligencia, víz alatti zaj, cetfélék akusztikus kommunikációja, víz alatti hangok osztályozása, neurális hálózat (NN)

1. BEVEZETÉS

A tengeri környezet egy összetett ökoszisztéma, amelyet a tengeri élővilág széles skáláját tart fenn, beleértve a cetféléket, például a bálnákat, delfineket. Ezek a tengeri emlősök kommunikációjuk, navigációjuk, táplálékszerzésük és egyéb alapvető viselkedésük során nagymértékben támaszkodnak a hangokra. A tenger akusztikus hangzásvilága azonban gyorsan változik az emberi tevékenységek, elsősorban az olyan forrásokból származó víz alatti zajok miatt, mint az idegenforgalom, a hajózás, az építkezések és az ipari tevékenységek. Ez a jelenség jelentős hatással van a tengeri élővilágra, különösen a cetfélékre, mivel zavarja a kommunikációt és a hatékony navigáció képességét. A cetfélék kifinomult hangadásukról ismertek, amely döntő szerepet játszik a társas interakciókban, a párzási rituálékban és az általános túlélésben. Ezek az állatok számos hangot adnak ki, a kattogástól és füttyentéstől kezdve az összetett énekekig, amelyek hatalmas távolságokat képesek megtenni a víz alatt. A cetfélék hangadásának tanulmányozásával a kutatók betekintést nyerhetnek viselkedésükbe, populációdinamikájukba és a környezeti változásokra adott válaszaikba. Ezen túlmenően a hangok megértése kulcsfontosságú a víz alatti zajcetpopulációkra gyakorolt hatásának felméréséhez. A víz alatti zajok és a cetfélékre gyakorolt hatásuk tanulmányozása az utóbbi években egyre nagyobb hangsúlyt kapott, ami a tengeri ökoszisztémák megőrzése iránti növekvő aggodalmat tükrözi. A kutatók különböző technikák kombinációját alkalmazzák a cetfélék hangadásának és a víz alatti zajra adott reakcióiknak a vizsgálatára. A víz alatti zaj és a cetfélék hangadásának tanulmányozása kiemelkedő jelentőségű a tengeri élővilág és akusztikus környezetük közötti bonyolult kapcsolatok

megértéséhez. Mivel az antropogén eredetű zaj egyre növekszik, e dinamika megértése kulcsfontosságúvá válik a cetpopulációk megőrzése és kezelése, valamint a tengeri ökoszisztémák általános egészsége szempontjából [1].

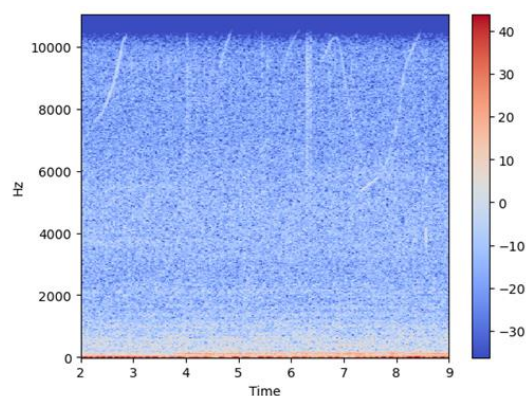
2. ANYAGOK ÉS MÓDSZEREK

Az Adriai-tengeren végzett terepmunka során hidrofont {Aquarian AS-1 típus, 1 Hz - 100 kHz \pm 2dB linearitás, -208 dBV to 1 μ Pa (40 μ V/Pascal) érzékenység} és egy Focusrite USB hangdigitalizáló modult (192 kHz mintavételi frekvencia, 24 bites felbontás) használtunk a víz alatti hangok "Wave" formátumú fájlok (.wav , PCM 24 bites kódolás, mintavételi frekvencia 192 kHz) rögzítésére. Ezeket a terepmunkám során rögzített fájlokat (időtartam: teljes idő 55 perc rövid hangklipek) használtam az NN-alapú modell tesztelésére az osztályozáshoz és a felismeréshez. Az adatgyűjtési folyamat során nagy mennyiségű víz alatti hangfelvétel keletkezik, amelyek emberi feldolgozására igen időigényes, és a munka monoton jellege miatt nagy az emberi hiba és az emberi figyelmetlenség lehetősége. Ezeket a szempontokat figyelembe véve felmerült a mesterséges intelligencia alkalmazásának lehetősége a hangfelvételek utólagos feldolgozására, a hangok osztályozására, a cetfélék és delfinek esetében a fajfelismerés szintjéig, a hajóforgalom zaja esetében pedig a hajótípus-felismerés szintjéig. Az osztályozáshoz a gépi tanulás neurális hálózati technikáját választottuk, mivel a hálózat automatikus tanulással tanulja meg a szűrők ("kernelek") optimalizálását. [2][3] [4] [5] [6] [7] [8] [9] [10] [11] [12]



1. ábra. Adatfeldolgozás gépi tanulással

A hangfájlok előfeldolgozásához, a jellemzők kinyeréséhez és a modellek létrehozásához Python programozási nyelvi csomagokat használtunk, a hangfájlok feldolgozásához a Librosa [13], az NN-modell létrehozásához és alkalmazásához, az osztályozáshoz a Keras [14], SciKitLearn [15] és TensorFlow [16] Python csomagokat, a vizualizációhoz a Matplotlib [17], a tensor- és tömbműveletekhez a Numpy [18], a statisztikai és numerikus számításokhoz a Pandas [19], a csv adatfájlok kezeléséhez pedig a Python csomagot. Az NN (neurális hálózat) modellünket korábban egy ember által osztályozott adatbázis segítségével képeztük ki. Az adatbázis, amelyet az NN gépi tanulási algoritmus betanításához használtunk, a Watkins Marine Mammal Sound Database, egy olyan forrás, amely több mint 60 tengeri emlősfaj mintegy 2000 egyedi hangfelvételét tartalmazza, több mint 15 000 annotált digitális hangfelvétellel, és több mint 70 év gyűjtőmunkájának gyümölcse. [20] A neurális hálózatok nem a nyers adatokat (például szövegfájlokat, hangfájlokat vagy CSV-fájlokat), hanem vektorizált és szabványosított reprezentációkat dolgoznak fel. Emiatt a .WAV hangfájlokat .CSV fájlkká ("vesszővel elválasztott értékek" fájl) alakítottuk át. A CSV-adatok elemzéséhez a numerikus jellemzőket lebegőpontos tenzorokká kellett átalakítani, a kategorikus jellemzőket pedig indexelni és egészértékű tenzorokká alakítani. Ezután minden jellemzőt normalizáltunk egy tipikus nulla átlaghoz és egy egyes szóráshoz. A "modell" egy rétegekből álló irányított aciklikus gráf. A modell úgy is elképzelhető, mint egy "nagyobb réteg", amely több alrétegből áll, és amelyet az adatokkal való érintkezés és találkozás révén lehet betanítani. Ezekhez a számításokhoz a Keras és a Tensorflow Python csomagok teljes támogatást nyújtanak, és képesek párhuzamosan futni több processzoron, akár grafikus feldolgozó egységeken (Nvidia GPU) is, nagyon jól skálázható végrehajtási teljesítmény mellett.



2. ábra. Egy hangfájl feldolgozása és a jellemzők kinyerése a Librosa Python csomag segítségével

A 2. ábra egy példát mutat az Adriai-tenger zajos víz alatti környezetében készített hangfájlból Librosa függvények segítségével létrehozott többdimenziós tömbökre. A hangfájlokból a Librosa függvények segítségével kinyert jellemzőket használjuk fel. [28] [29].

A nyílt forráskódú szoftverek és a szabadon hozzáférhető adattárak átalakító megközelítést jelentenek a mai tudományos és technológiai környezetben. Az olyan platformok megjelenésével, mint a GitHub, valamint az olyan licencmechanizmusokkal, mint a GNU General Public License (GPL), az együttműködés és az átláthatóság egyre szélesebb horizontja nyílik meg. Ez a tanulmány a nyílt forráskódú programok, a GNU licenckel alapjául szolgáló elvek és a GitHub szerepét vizsgálja a közösség által irányított projektfejlesztés és az adatok terjesztésének elősegítésében[28] [29].

Az ilyen platformok nemcsak a kódok terjesztésének infrastruktúráját biztosítják, hanem az adatok demokratizálását is elősegítik. Az előnyök sokrétűek: felgyorsítják a kutatást, biztosítják a reprodukálhatóságot és elősegítik a globális együttműködést földrajzi vagy pénzügyi korlátok nélkül.

A Watkins Marine Mammal Sound Database [20] 15 567 wav fájljából (10,3 GB) ezeket a jellemzőket egyenként kivonatoltuk és csv fájlban tároltuk. Az így kapott csv-fájl a jellemzőket normalizált formában tartalmazza. A létrehozott címkézett adatbázis (csv fájl) segítségével betanítottuk az NN modellünket, ahogy a 3. ábrán látható, 189.229 paraméter állt rendelkezésre a modellben

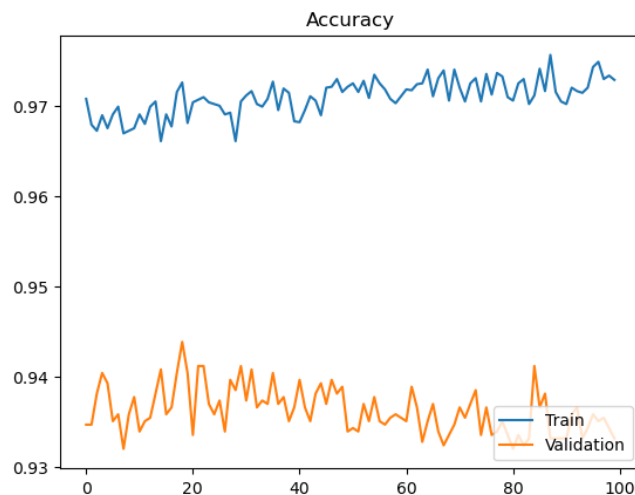
```
print(model.summary())
```

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 512)	13824
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 256)	131328
dropout_1 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 128)	32896
dropout_2 (Dropout)	(None, 128)	0
dense_3 (Dense)	(None, 64)	8256
dropout_3 (Dropout)	(None, 64)	0
dense_4 (Dense)	(None, 45)	2925
Total params: 189,229		
Trainable params: 189,229		
Non-trainable params: 0		

3. ábra NN modell létrehozása

A kivont jellemzők "illeszkedésének jóságát" és a modell képzésére való alkalmasságát a validálással teszteltük (lásd a 4. ábrát).

```
acc_train_curve = model_history.history["accuracy"]
acc_val_curve = model_history.history["val_accuracy"]
plt.plot(acc_train_curve, label = "Train")
plt.plot(acc_val_curve, label = "Validation")
plt.legend(loc = 'lower right')
plt.title("Accuracy")
plt.show()
```



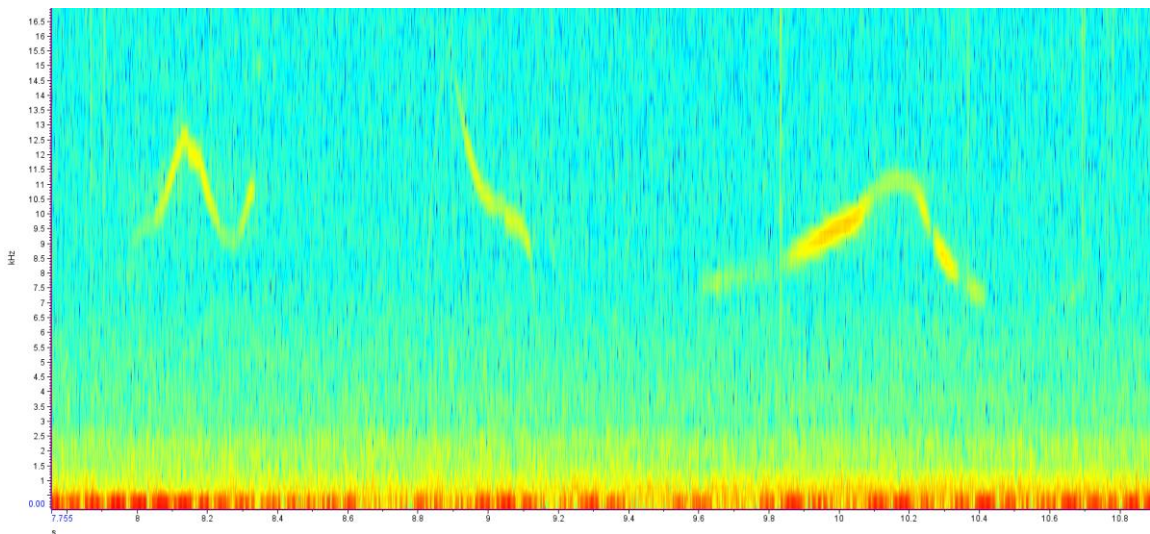
4. ábra. A modell pontossága.

3. EREDMÉNYEK

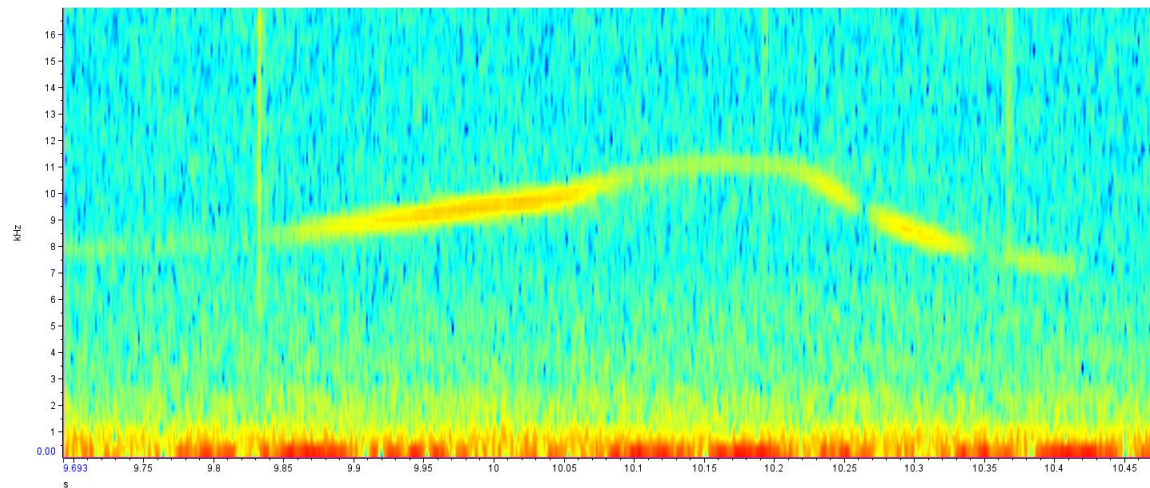
A mintaadatbázison betanított modell maximális pontossága 93,31% volt a modell kiértékelésének eredményeként.

Ezzel gyakorlatilag elkészült az NN-modellünk, amely alkalmas új, akár nagyméretű, a modell számára ismeretlen hangfájlok osztályozására, azaz a bennük található cetfélék hangjainak felismerésére és fajmeghatározására, automatikus, emberi beavatkozás nélküli osztályozására.

A 5. és 6. ábra példát mutat a delfinhangokat tartalmazó hangfájl spektrumára, amelyek a kedvtelési célú hajók által keltett nagy zajjal rendelkeznek.



5. ábra. Az általunk az Adriai-tengeren rögzített delfinhangok spektrumja, a kedvtelési célú hajók nagy zajával



6. ábra. Az általunk az Adriai-tengeren rögzített, erősen modulált delfinhang spektrumja, a kedvtelési célú hajók nagy zajával

Az Adriai-tengeren, Veli Losinj (Jadransko more, Adriai-tenger, WGS84: 44.53417 N, 14.52832 E, 13.09.2022 9:28) térségében rögzített hidrofon hangfelvételek az 5. és 6. ábrán láthatók, ahol a delfinek kattogása (click) és sípolása (whistle), valamint a hajóforgalom zaja látható.

A gazdag biológiai sokféleségéről és történelmi jelentőségéről ismert Adriai-tenger az elmúlt évtizedekben a tengeri turizmus óriása növekedését élte és éli át. Így a turizmusra épülő hajók által kibocsátott víz alatti zaj jelentős hatással lehet a régió tengeri élővilágára. Ez a tanulmány az Adriai-tengeren

közlekedő, turizmusra épülő hajók akusztikai lábnyomát és annak lehetséges ökológiai következményeit igyekszik megvilágítani.

A víz alatti zajszintek szezonális ingadozása a turisztikai főszezonnak megfelelő. A nagyobb kikötőkhöz és a történelmi jelentőségű helyszínekhez közelebb eső régiókban magasabb zajamplitúdót mértek. A spektrumanalízis 20-3000 Hz közötti domináns frekvenciákat mutatott ki, esetenként a fedélzeti szórakoztató rendszerekhez vagy a hajógépekhez kapcsolódó magasabb frekvenciájú zajforrásokkal. Figyelemre méltó, hogy a megnövekedett zajszintű területek több cetfaj, köztük a palackorrú delfin (*Tursiops truncatus*) élőhelyeivel és vándorlási útvonalalaival fedik egymást.

Az Adriai-tengeren a turizmusra épülő hajók által okozott megnövekedett zajszint jelentős aggodalomra adhat okot a tengeri élővilág számára. Az ilyen zajnak való krónikus kitettség viselkedésbeli változásokhoz, stresszhez és a kommunikáció elfedéséhez vezethet a ceteknél és más, akusztikailag érzékeny fajoknál.

A további feladatok közé tartozik a teljes hidrofon hangrögzítő rendszer teljes automatikus, terepen elvégezhető kalibrációjának kifejlesztése (mértékegység dB re 1 uPa.), a víz alatti hangok mérésére. [29]

A víz alatti zaj tengeri vagy édesvízi tavak biológiai sokféleségére gyakorolt hatásának felmérése fontos az érzékeny tengeri és édesvízi fajok hatékony védelme, valamint a tengeri, part menti és édesvízi ökoszisztémák fenntartható használatának elősegítése érdekében. [22, 23, 24, 25, 26, 27] A terepi adatgyűjtés nagy mennyiségű adatot generál, amelyek feldolgozása, elemzése és osztályozása mesterséges neurális hálózatok segítségével automatizálható és jelentősen felgyorsítható. A delfinek zajos környezetben történő kommunikációjának tanulmányozása azt mutatja, hogy a delfinek szonárrendszere, amely "kattogásnak" nevezett hangimpulzusokat használ a környezetük pásztázására, a hasonló méretű és frekvenciájú emberi képalkotó szonárokhoz képest kiváló alakmegkülönböztető képességgel rendelkezik. [27] Az előzetes információk beépítése a jelfeldolgozásba, különösen az alakzatok ritkaságára és gyakoriságára vonatkozó információk, a hagyományos jelfeldolgozáshoz képest a visszavert hangok egyértelműbb értelmezését eredményezi. A mesterséges neurális hálózatokat (a delfin szonár utánzó) egy biomimetikus szonárrendszer kifejlesztésére használhatjuk, amely egy kompakt, nagyobb felbontású, zajos környezetben is jól működő szonár képalkotó technológia felé nyitja meg az utat, a különböző lehetséges ipari felhasználásokkal. [27]

KÖSZÖNETNYILVÁNÍTÁS

Szeretnénk köszönetet nyilvánítani az ION-technik Kft. és az ION Alkalmazott Kutatási NP Kft. támogatásáért.

IRODALMI HIVATKOZÁSOK

- [1] "Soundscapes in the North Adriatic Sea and their impact on marine biological resources," Retrieved from <https://www.italy-croatia.eu/web/soundscape>
- [2] M.V. Valueva, et al., "Application of the residue number system to reduce hardware costs of the convolutional neural network implementation, Convolutional neural networks are a promising tool for solving the problem of pattern recognition.," *Mathematics and Computers in Simulation*. Elsevier BV. 177: 232-243., 2020.
- [3] Z. Wei, "Shift-invariant pattern recognition neural network and its optical architecture,," *Proceedings of Annual Conference of the Japan Society of Applied Physics.*, 1988.
- [4] Z. Wei, "Parallel distributed processing model with local space-invariant interconnections and its optical architecture". *Applied Optics*. 29 (32): 4790-7. 1990.
- [5] A. Mouton, et al., "Artificial Intelligence Research.," *Communications in Computer and Information Science*. Cham: Springer International Publishing. 1342: 267-281., 2020.
- [6] Van Den Oord, et al., "Deep content-based music recommendation.," *Curran Associates, Inc.* pp. 2643-2651., 2013.
- [7] Collobert, et al. "A unified architecture for natural language processing: deep neural networks with multitask learning.," *Proceedings of the 25th international conference on machine learning*. ICML '08. pp. 160-167., 2008.
- [8] Avilov, et al., "Deep Learning Techniques to Improve Intraoperative Awareness Detection from Electroencephalographic Signals,," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Montreal, QC, Canada: IEEE. 2020: 142-145.
- [9] Tsantekidis, et al. "Forecasting Stock Prices from the Limit Order Book Using Convolutional Neural Networks,," 2017 IEEE 19th Conference on Business Informatics (CBI). Thessaloniki, Greece: IEEE: 7-12.
- [10] K. Fukushima, "Neocognitron.," *Scholarpedia*. 2 (1): 1717., 2007.
- [11] Fukushima, Kunihiko, "Neocognitron: "A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position,," *Biological Cybernetics*. 36 (4): 193-202., 1980.

- [12] Matusugu, et al., "Subject independent facial expression recognition with robust face detection using a convolutional neural network," *Neural Networks*. 16 (5): 555-559., 2003.
- [13] Mcfee, et al, "librosa: Audio and music signal analysis in python.," In *Proceedings of the 14th python in science conference*, pp. 18-25. 2015. <https://zenodo.org/badge/latestdoi/6309729>
- [14] F. Chollet, et al., "keras. gitHub.," Retrieved from <https://github.com/fchollet/keras>.,2015.
- [15] F. Pedregosa, et al., "scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12(Oct), 2825-2830., 2011.
- [16] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, R. Jozefowicz, Y. Jia, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, M. Schuster, R. Monga, S. Moore, D. Murray, C. Olah, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng., "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. software available from tensorflow.org. <https://doi.org/10.5281/zenodo.4724125>
- [17] J. D. Hunter, "Matplotlib: A 2D Graphics Environment", *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90-95, 2007. <https://doi.org/10.5281/zenodo.592536>
- [18] Harris, C.R., Millman, K.J., Van Der Walt, S.J. Et Al. *array programming with NumPy*. *Nature* 585, 357-362 (2020). DOI: 10.1038/s41586-020-2649-2.
- [19] Mckinney, "Data structures for statistical computing in python," *Proceedings of the 9th Python in Science Conference*, Volume 445, 2010. Pandas. <https://doi.org/10.5281/zenodo.3509134>
- [20] J. Shrey, "Best of Watkins Marine Mammal Sound Database," Retrieved from <https://www.kaggle.com/datasets/shreyj1729/best-of-watkins-marine-mammal-sound-database>
- [21] M Picciulin, M. Bolgan, N Rako-Gospić, A Petrizzo, M. Radulović, R. Falkner,"A Fish and Dolphin Biophony in the Boat Noise-Dominated Soundscape of the Cres-Lošinj Archipelago (Croatia). *Journal of Marine Science and Engineering*, 10(2), p.300., 2022.
- [22] M. Picciulin, E Armelloni, R Falkner, N Rako-Gospić, , M. Radulović, G. Pleslić, S. Muslim, H. Mihanović, T. Gaggero,"characterization of the underwater noise produced by recreational and small fishing boats (< 14 m) in the shallow-water of the Cres-Lošinj," *Natura 2000 SCI. Marine Pollution Bulletin*, 183, p.114050., 2022.
- [23] A.R. Luís, et al, "Vocal universals and geographic variations in the acoustic repertoire of the common bottlenose dolphin.," *Scientific reports*, 11(1), pp.1-9., 2021.
- [24] N. Gospić, M Picciulin," Changes in whistle structure of resident bottlenose dolphins in relation to underwater noise and boat traffic. *Marine Pollution Bulletin* 105, 193-8., 2016.
- [25] N. Rako-Gospić, N. Radulović, M. Vučur, T. Pleslić, G. Holcer, P. Mackelworth, "Factor associated variations in the home range of a resident Adriatic common bottlenose dolphin population.," *Marine Pollution Bulletin*. 2017.
- [26] Rako N., et al., "Leisure boating noise as a trigger for the displacement of the bottlenose dolphins of the Cres-Lošinj archipelago (northern Adriatic Sea, Croatia) *Marine Pollution Bulletin* 68, 77-84., 2013.
- [27] H. Vishnu, H., M. Hoffmann-Kuhnt, M. Chitre, A. Ho, E. Matrai,"A dolphin-inspired compact sonar for underwater acoustic imaging.," *communications engineering* 1, 10 <https://doi.org/10.1038/s44172-022-00010-x>
- [28] GitHub: Analyze and classify sounds with AI. Retrieved from Github: <https://github.com/ovh/ai-training-examples/blob/main/notebooks/audio/audio-classification/notebook-marine-sound-classification.ipynb>
- [29] GitHub: A. Aradi, SoundScape and Dolphins Retrieved from Github: https://github.com/capnA2XY/SoundScape_and_Dolphins/tree/main