JASA ARTICLE

Introduction to the special issue on machine learning in acoustics[a]

Zoi-Heleni Michalopoulou,[1,b] Peter Gerstoft,[2,c] Bozena Kostek,[3,d] and Marie A. Roch[4,e]

[1]*Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, New Jersey 07102, USA*

[2]*Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093, USA*

[3]*Faculty of Electronics, Telecommunications and Informatics, Audio Acoustics Laboratory, Gdansk University of Technology (GUT), Gdansk, Poland*

[4]*Department of Computer Science, San Diego State University, San Diego, California 92182-7720, USA*

ABSTRACT:

The use of machine learning (ML) in acoustics has received much attention in the last decade. ML is unique in that it can be applied to all areas of acoustics. ML has transformative potentials as it can extract statistically based new information about events observed in acoustic data. Acoustic data provide scientific and engineering insight ranging from biology and communications to ocean and Earth science. This special issue included 61 papers, illustrating the very diverse applications of ML in acoustics.

https://doi.org/10.1121/10.0006783

(Received 21 September 2021; accepted 1 October 2021; published online 29 October 2021)

[Editor: James F. Lynch] Pages: 3204–3210

I. INTRODUCTION

When we started our Call for Papers for a Special Issue on "Machine Learning in Acoustics" in the *Journal of the Acoustical Society of America*, our ambition was to invite papers in which machine learning was applied to all acoustics areas. They were listed, but not limited to, as follows:

- Music and synthesis analysis
- Music sentiment analysis
- Music perception
- Intelligent music recognition
- Musical source separation
- Singing analysis and voice quality evaluation
- Expressivity in music
- Bioacoustics
- Soundscapes
- Hearing and hearing aids
- Speech, language, and emotion recognition
- Speech recognition
- Emotion in speech
- Speech perception
- Expressivity in speech
- Intelligent speech processing
- Multimedia speech processing
- Classification from active acoustics
- Acoustic source localisation
- Acoustic field prediction in ocean acoustics
- Acoustical oceanography.

This is because—through the last decades—we had seen that machine learning had found its rightful place in acoustics, especially when a particular area needed a novel approach to challenging "old" complex problems. Years ago, Zadeh[1,2] defined soft computing, a group of machine learning techniques that permit the input data and the problem description to be imprecise. This involved several computing methods, i.e., artificial neural networks, fuzzy logic, and probabilistic reasoning, e.g., rough sets, etc., oriented towards human-like reasoning. All these methods concurred with a statistical approach to justify the results obtained. One may wonder what could be the cause of the relatively rapid recent surge of machine learning methods and their progress, often referred to as deep learning. It includes Bayes networks, neural networks, and adversarial networks, to name a few. Of course, there is no one answer to this question; there are several apparent causes that have a noticeable effect on the advances of machine learning algorithms and applications, i.e., the rapid growth of database resources that facilitate gathering massive amounts of data and often sharing them among scientific communities, faster processors, using graphic cards to process data and signals, networking the resources, faster networks, and open access to research.

In this special issue, there is a variety of such machine learning techniques and their application to many acoustics areas; however, the interest of researchers is not distributed evenly. There are many papers related to underwater acoustics, including marine mammal sound analyses, noise, echolocation, auditory scene analysis, and another area of the highest interest is speech processing. Figure 1 shows the frequency of occurrences of subjects to visualise the distribution of papers in which machine learning methods were used in acoustics.

[a]This paper is part of a special issue on Machine Learning in Acoustics.

[b]ORCID: 0000-0001-8153-7851.

[c]Electronic mail: gerstoft@ucsd.edu, ORCID: 0000-0002-0471-062X.

[d]ORCID: 0000-0001-6288-2908.

[e]ORCID: 0000-0002-0687-2059.

https://doi.org/10.1121/10.0006783

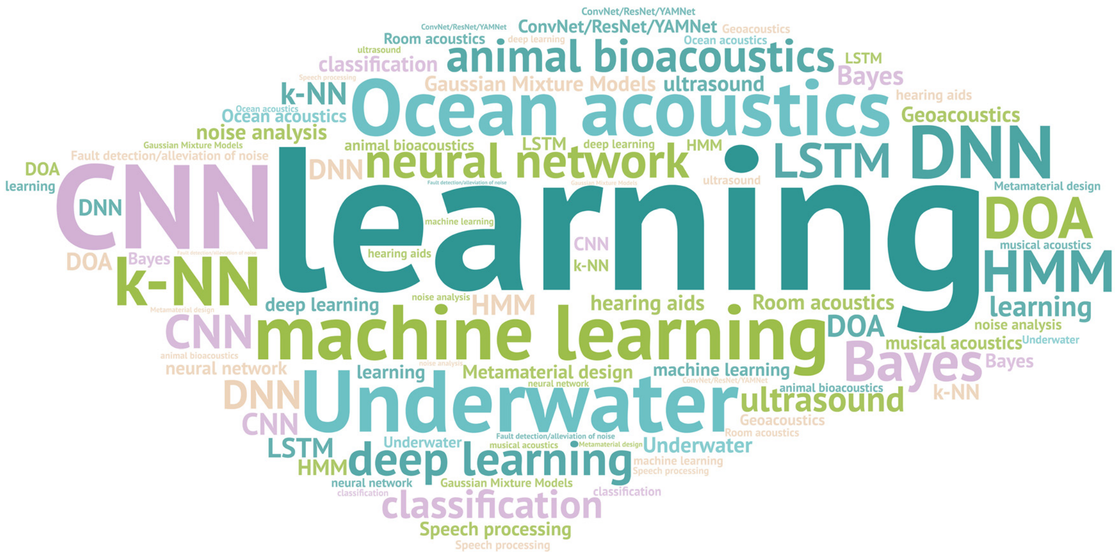

FIG. 1. (Color online) Frequency of occurrences of a particular subject in this issue related to acoustic areas in which machine learning was applied.

Acoustics Today has a companion paper to this issue, "How machine learning contributes to solve acoustical problems?," prepared by our group of guest editors.[3] We encourage readers that are unfamiliar with machine learning to examine this article; it may be a smooth introduction for new adepts in acoustics searching to understand the ideas behind machine learning. For an introduction to the mathematical details of machine learning techniques, please see Ref. 4. We hope that the papers in this special issue will be of interest to the readers, particularly given that mature up-to-date machine learning methods are applied to acoustics.

II. CONTRIBUTED PAPERS

A. Ocean acoustics

Ocean acoustics is an area that presents us with rich opportunities for the development and application of machine learning techniques to a variety of complex problems. Seabed parameter characterisation, often referred to as geoacoustic inversion, is a topic that has attracted significant interest in the past. Along these lines, in Ref. 5, Frederick *et al.* implemented machine learning approaches including convolutional neural networks (CNNs) for sediment classification from acoustic fields and showed that, in the presence of uncertainty, machine learning approaches outperform conventional matched-field processing (MFP). Similarly, Smaragdakis and Taroudakis[6] implemented a Hidden Markov model approach to extract features from acoustic signals in the ocean coupled with genetic algorithms and developed a reliable method for geoacoustic inversion. Shen *et al.*[7] employed machine learning for geoacoustic inversion as well. Results therein indicate that their technique, involving radial basis function networks, leads to higher sensitivity in the inversion, with respect to some of the unknown sediment parameters in comparison to that of conventional methods.

Source localisation is another problem that has been the focus of a significant body of research in ocean acoustics and is now being extensively addressed with machine learning. Liu *et al.*[8] applied CNNs and multi-task learning to the problem of source localisation in an underwater environment. They found that CNNs are more robust to environmental uncertainty in comparison to conventional MFP. Chen and Schmidt[9] also used CNNs for source localisation in the ocean. Similarly, to other studies, they demonstrated with both synthetic and real data that the implemented networks outperform conventional MFP in the presence of environmental uncertainty. Ferguson[10] used cepstrum-based and correlation-based feature vectors to train CNNs for the localisation of a transiting motorboat. He showed that the combined cepstrum–cross correlation CNN provides superior source localisation performance in comparison to CNNs where only cepstrum or only correlation is used as features. Wang *et al.*,[11] conducted source depth estimation in the ocean using a combination of CNNs and conventional beamforming, while also employing transfer learning. Results demonstrated the superiority of the method over conventional CNNs. Neilsen *et al.*[12] performed source depth estimation in an oceanic environment–simultaneously with seabed classification–using CNNs as well. They investigated the effect of mismatch between training and testing patterns on the estimation/classification results. Along the same lines, Van Komen *et al.*[13] carried out source localisation and sediment characterisation by applying CNNs to spectrograms calculated from surface ship sound signals.

Yoon *et al.*[14] performed source depth estimation in the ocean using residual neural networks and showed robustness of the approach with respect to environmental uncertainty and receiving sensor position. A deep transfer learning method was proposed by Cao *et al.*[15] for direction of arrival estimation using a single-vector sensor. Performance analysis showed that the approach outperforms conventional

CNNs. Data recorded by vector sensors were also used in Ref. 16. Whitaker *et al.*[16] used recurrent neural networks, both shallow and deep, for direction of arrival estimation in an oceanic medium using a vector sensor. The network performance was superior to that of conventional estimation. Deep networks performed better than shallow ones.

Other applications in ocean acoustics include acoustic mine detection and underwater communications. Brandes *et al.*[17] implemented an environmentally adaptive approach coupled with a segmentation constrained neural network for effective automated mine-like object recognition. McCarthy *et al.*[18] applied a model-agnostic geometric feature representation based on braid theory to track diverse channel phenomena and improve channel estimation of shallow water acoustic communications. Moreover, these techniques proved valuable in feature extraction for machine learning using feed forward neural networks in estimation and determining the number of reflector activity tracked by braids may undergo. Zhang *et al.*[19] applied a meta-learning approach to orthogonal frequency division multiplexing for underwater acoustic communications and showed that the method is superior to conventional deep neural networks, frequently used in such tasks.

B. Animal bioacoustics

Observations of animal life and detection and classification of biological signals also present us with a fertile ground for the development and application of diverse machine learning approaches. Methods range from detecting species or populations to extracting detailed characteristics of calls. Many of these detection and classification techniques are first steps of other processes, such as mitigation, density estimation, or monitoring behaviour.

Lee and Staneva[20] developed an unsupervised machine learning approach based on matrix decomposition for extracting information from data collected from ecosystems using echosounders; this approach facilitated acoustics-based biological observation in the ocean. Ozanich *et al.*[21] applied deep embedded clustering and Gaussian mixture models to distinguish fish sounds from whale vocalizations. They found that these methods are superior to conventional clustering. Cotter *et al.*[22] implemented a k-nearest-neighbours approach using scattering models appropriate for echosounder measurements and employed it to group organism target spectra measured in the mesopelagic near the New England continental shelf break. Gruden and White[23] showed how marine mammal vocalizations can be extracted efficiently and effectively from recorded data using machine learning techniques based on probability hypothesis density filters. Gruden *et al.*[24] also proposed a multi-target tracking method to automate time difference of arrival tracking, based on the Gaussian mixture probability hypothesis density filter and including multiple sources, source appearance and disappearance, missed detections, and false alarms. It was shown that by using an extended measurement model and combining measurements from broadband echolocation clicks and narrowband whistles, more information can be extracted from acoustic encounters. The method was tested on false killer whale recordings.

On the problem of identifying marine mammal vocalizations, Rasmussen and Sirovic[25] developed CNN-based detectors with high precision and recall for highly variable whale calls. Padovese *et al.*[26] investigated the benefit of augmenting training datasets with synthetically generated samples when training a deep neural network for the classification of North Atlantic right whale upcalls. Two augmentation techniques, SpecAugment and Mixup, were implemented and were shown to improve call classification. Denoising CNNs and autoencoders were applied to spectrograms of noise-corrupted audio signals by Vickers *et al.*[27] The denoised signals were then included in training, leading to high accuracy in detection/classification of whale vocalizations. Schall *et al.*[28] showed that accurate discriminant models, based on combinations of acoustic metrics and supervised neural networks, provide an automated solution for fast and highly reproducible identification and comparison of vocalisation types in humpback whale populations. Roch *et al.*[29] used feed forward neural networks to develop a time-domain whale echolocation click detector. The approach was capable of finding echolocation bouts that were missed by human analysts. Zhong *et al.*[30] used acoustic signatures to detect, classify, and count the calls of four acoustic populations of blue whales. The goal was that the conservation status of each population is better assessed. Siamese neural networks were used for the task, which were found to outperform CNNs.

Along similar lines to marine life–related studies, Morfi *et al.*[31] created deep perceptual embeddings for bird sounds using triplet networks. They used multidimensional scaling pretraining, attention pooling, and a triplet mining scheme. They then demonstrated the feasibility of the method to develop perceptual models for a wide range of data based on behavioural judgements, assisting in the understanding of how animals perceive sounds.

Classification of foliage was also approached with machine learning methods. Kuc[32] employed spectrograms echoes from foliage targets as input to artificial neural networks for foliage target classification. Kuc showed that classification performance using spectrograms is significantly superior to performance when echo envelopes are employed.

C. Metamaterial design

The study and design of metamaterials is yet another field that offers rich opportunities for the use of machine learning. Ciaburro and Iannace[33] collected metamaterial sound absorption coefficient measurements. These measurements were employed to train artificial neural networks. When tested, the networks provided acoustic absorption coefficient estimates of the metamaterial, which are very similar to the measured values. Gurbuz *et al.*[34] used adversarial neural networks to obtain insights into the design of acoustic metamaterials. Shah et al.[35] also used machine learning towards metamaterial design. Specifically, reinforcement learning was employed towards acoustic material design with cloaking as the ultimate goal.

Material properties were studied by Stender et al.[36] as well. A data-driven reverse engineering approach was used to identify factors from absorption coefficient spectra of sound absorbing materials. The analysis of a neural network identified important features in absorption coefficient spectra. The results indicated the amount to which different factors affect the absorption coefficient measurements and may help to better understand how different manufacturing technologies or mounting approaches affect the absorption coefficient at which frequencies.

D. Speech processing

The processing of speech–for enhancement, recognition, synthesis, or emotion perception to name a few directions–has been tremendously advanced with the development of new and powerful machine learning approaches. These techniques are disrupting standard methods of human-machine interaction and have a wide range of applications, including the quickly growing Internet of things.

Shankar et al.[37] implemented a two-microphone speech enhancement framework relying on recurrent neural networks. The approach improved speech quality and intelligibility in noisy environments. In parallel, Chinen et al.[38] investigated a multidimensional mapping function using deep lattice networks for speech quality estimation in the presence of a variety of distortions. The approach outperformed speech quality estimation with conventional mapping functions and facilitated uncertainty quantification. Morgan et al.[39] analysed the performance of several neural network architectures, including CNNs for predicting speech emotion, aiming at the prediction of emergent leadership among other group metrics.

Liu et al.[40] used spectrograms of speech signal segments, along with the actual waveforms, and successfully applied FaceNet to them for emotion recognition, whereas Mahmud et al.[41] implemented support vector machine models for the assessment of how well listeners' speech categorisation can be decoded via whole-brain and hemisphere-specific responses. Zhang et al.[42] developed a high resolution direction-of-arrival approach based on deep neural networks for multiple speech source localisation using a small scale array. The new approach outperformed conventional beamforming techniques.

Riad et al.[43] proposed the use of a parametrised neural network layer, computing specific spectro-temporal modulations based on Gabor filters. The approach exhibited excellent performance in speech activity detection, whereas it was also successful and comparable to state-of-the art approaches in speaker verification, urban sound classification, and zebra finch call type classification. Piotrowska et al.[44] employed k-nearest neighbours, the naive Bayes method, long–short term memory, and CNNs towards automated evaluation of pronunciation focussed on a particular phonological feature. Korvel et al.[45] identified a way of highlighting the acoustic differences between consonant phonemes of the Polish and Lithuanian languages. Similarity matrices were employed based on speech acoustic parameters combined with a CNN. The performance of the similarity matrix approach demonstrated its superiority over other techniques. The work in Ref. 46 showed that machine learning techniques are very successful at classifying the Russian fricatives [f], [s], and [ʃ] using a small set of acoustic cues. Classifiers based on decision trees, random forests, support vector machines, and neural networks were implemented to distinguish between the three fricatives. The results demonstrated successful classification.

Tsipas et al.[47] introduced and evaluated an audio-driven, multimodal approach for speaker diarization in multimedia content using semi-supervised clustering of audio-visual embeddings, generated using deep learning techniques.

Smalt et al.[48] explored using visual information from a photograph of a hearing protection device (HPD) inserted into the ear to estimate hearing protector attenuation. Using a deep neural network, high classification accuracy was achieved in predicting if the HPD fit was greater or smaller than the median measured attenuation.

E. Fault detection

In this special issue, studies are also presented on fault detection and also alleviation of noise. Alavijeh et al.[49] investigated the applicability of machine learning to the automation of the ultrasonic inspection of polyethelene pipe butt-fusion joints. CNNs were found to be effective tools for the detection of defects. Along the same lines of fault identification, in Ref. 50, a support vector machine was used to evaluate feature importance for squeak and rattle identification. Similarly, Teja et al.[51] worked on the identification of sloshing noises in fuel tanks using CNNs with the creation of quieter tanks being the goal of the study. The identification accuracy of the proposed CNN is about 94%. Mei et al.[52] studied a robot-assisted ultrasonic testing system using the track-scan imaging method for improving the detecting coverage and contrast of ultrasonic images. They proposed a visual geometry deep learning network to optimise the reconstructed ultrasonic images. The results indicate that the proposed method improves the resolution of reconstructed ultrasonic images without sacrificing efficiency.

F. Room acoustics

Liu et al.[53] approached sound source localisation in noisy and reverberant rooms using microphone arrays as a classification task. Simulation and real-world experimental results demonstrated that the proposed deep learning assisted approach can achieve higher spatial resolution and is superior to other state-of-the-art techniques.

Foy et al.[54] introduced a new supervised learning approach to estimate the mean absorption coefficients from a room impulse response (RIR) regarding building acoustics and the acoustic diagnosis of an existing room. The RIR-to-absorption mapping was learned by regression on a simulated dataset using artificial neural networks.

Shalev *et al.*[55] introduced an extension of the image method for generating room impulse responses in a structure with more than a single confined space. The proposed approach can generate, in an efficient manner, a large number of environmental examples for a structure impulse response, required by current deep-learning methods for many tasks.

De Salvio *et al.*[58] looked into the problem of noisy environments in offices and developed Machine Learning approaches to identify human and mechanical noise sources during working hours. Clustering techniques were implemented to obtain information on the number of sources, which were then labeled with the help of statistical and metrical features.

Tsokaktsidis *et al.*[59] studied noise in passenger cars. They used Artificial Neural Networks to predict interior noise in vehicles for different operational conditions. Their approach was accurate and cost-effective compared to standard practices of measuring transfer functions and numerically modeling the noise.

G. Musical acoustics

Colonel and Reiss[56] developed a method to retrieve the parameters used to create a multitrack mix using only raw tracks; the stereo mixdown is presented. This method is able to model linear time-invariant effects, such as gain, pan, equalisation, delay, and reverb. Pujol *et al.*[57] used a multiresolution deep learning approach that allows the encoding of information contained in unprocessed time-domain acoustic signals captured by microphone arrays. Results show that the BeamLearning approach outperforms the wideband MUSIC and steered response power-phase transform methods with respect to localisation accuracy and computational efficiency in the presence of noise and reverberation.

Hawley and Morrison[60] presents a CNN-based model for detecting and counting the vibration patterns from the electronic speckle pattern interferometry (ESPI) frames of steelpan vibration images. For gathering data, about 1200 human-annotated frames were crowdsourced from the Zooniverse Steelpan Vibration Project (ZSVP).[60] As an alternative approach, a much larger number of synthetic frames were generated, and the network was trained on both sets of data.

H. Environmental acoustic monitoring

Hart *et al.*[61] quantify the accuracy of three machine-learning models for long-range sound propagation, considering at the same time atmospheric turbulence. A synthetic dataset is generated by a parabolic equation model and is used for training and testing three machine-learning algorithms. The errors of these models with respect to an experimental long-range sound propagation dataset were studied. Gontier *et al.*[62] proposed a two-stage approach for environmental acoustic monitoring. In the self-supervised stage, they formulated a pretext task on unlabelled spectrograms from an acoustic sensor network. On the other hand, in the supervised stage, they formulated a downstream task of multilabel urban sound classification on synthetic scenes. They concluded that training set synthesis benefits monitoring performance more than self-supervised learning.

Chen *et al.*[63] proposed a long-term wavelet feature for acoustic scene classification (ASC) that captures discriminative long-term scene information. A data augmentation scheme was implemented that improved the generalisation of the ASC systems. In Ref. 64, two methods were presented that enabled the automated identification of aeroacoustic sources in sparse beamforming maps and the extraction of their corresponding spectra to overcome the manual definition of regions of interest. Both methods were found to be robust to statistical noise and predicted the source existence, location, and spatial probability estimation.

In Ref. 65, non-negative matrix factorisation was implemented to estimate the noise effects of wind turbines continuously, without interrupting their function. That facilitated the effective characterisation of the noise impact of wind farms.

I. Event detection and enhancement

Ekpezu *et al.*[66] showed that, using CNNs and long short-term memory, acoustic signals are effective for classifying and potentially detecting natural disasters.

Paul *et al.*[67] proposed a machine learning algorithm, based on a multilayer perceptron coupled with singular value decomposition, which was applied to several acoustical classification problems. The experiments quantified the extent to which closely related spectra can be distinguished. Shao *et al.*[68] investigated the performance of CNNs and transfer learning on thyroid tumour grade identification based on ultrasound images. They found that some implementations of transfer learning outperform CNNs in tumour grading.

ACKNOWLEDGMENTS

We would like to thank the reviewers for carefully assessing all papers submitted to this special issue. Many thanks are due to James F. Lynch, JASA Editor-in-Chief, Liz Bury, Senior Managing Editor, Kelly Quigley, and a team of copy editors for their help throughout the production process.

[1]L. A. Zadeh, "Fuzzy logic, neural networks, and soft computing," Commun. ACM **37**(3), 77–84 (1994).
[2]L. Zadeh, "From computing with numbers to computing with words. from manipulation of measurements to manipulation of perceptions," IEEE Trans. Circuits Syst. I: Fundamental Theory Appl. **46**(1), 105–119 (1999).
[3]M. A. Roch, P. Gerstoft, B. Kostek, and Z.-H. Michalopoulou, "How machine learning contributes to solve acoustical problems," Acoust. Today **17**(4) (2021).
[4]M. J. Bianco, P. Gerstoft, J. Traer, E. Ozanich, M. A. Roch, S. Gannot, and C.-A. Deledalle, "Machine learning in acoustics: Theory and applications," J. Acoust. Soc. Am. **146**(5), 3590–3628 (2019).

https://doi.org/10.1121/10.0006783

[5] C. Frederick, S. Villar, and Z.-H. Michalopoulou, "Seabed classification using physics-based modelling and machine learning," J. Acoust. Soc. Am. **148**(2), 859–872 (2020).

[6] C. Smaragdakis and M. I. Taroudakis, "Acoustic signal characterisation based on hidden Markov models with applications to geoacoustic inversions," J. Acoust. Soc. Am. **148**(4), 2337–2350 (2020).

[7] Y. Shen, X. Pan, Z. Zheng, and P. Gerstoft, "Matched-field geoacoustic inversion based on radial basis function neural network," J. Acoust. Soc. Am. **148**(5), 3279–3290 (2020).

[8] Y. Liu, H. Niu, and Z. Li, "A multi-task learning convolutional neural network for source localisation in deep ocean," J. Acoust. Soc. Am. **148**(2), 873–883 (2020).

[9] R. Chen and H. Schmidt, "Model-based convolutional neural network approach to underwater source-range estimation," J. Acoust. Soc. Am. **149**(1), 405–420 (2021).

[10] E. L. Ferguson, "Multitask convolutional neural network for acoustic localisation of a transiting broadband source using a hydrophone array," J. Acoust. Soc. Am. **150**(1), 248–256 (2021).

[11] W. Wang, Z. Wang, L. Su, T. Hu, Q. Ren, P. Gerstoft, and L. Ma, "Source depth estimation using spectral transformations and convolutional neural network in a deep-sea environment," J. Acoust. Soc. Am. **148**(6), 3633–3644 (2020).

[12] T. B. Neilsen, C. D. Escobar-Amado, M. C. Acree, W. S. Hodgkiss, D. F. Van Komen, D. P. Knobles, M. Badiey, and J. Castro-Correa, "Learning location and seabed type from a moving mid-frequency source," J. Acoust. Soc. Am. **149**(1), 692–705 (2021).

[13] D. F. Van Komen, T. B. Neilsen, D. B. Mortenson, M. C. Acree, D. P. Knobles, M. Badiey, and W. S. Hodgkiss, "Seabed type and source parameters predictions using ship spectrograms in convolutional neural networks," J. Acoust. Soc. Am. **149**(2), 1198–1210 (2021).

[14] S. Yoon, H. Yang, and W. Seong, "Deep learning-based high-frequency source depth estimation using a single sensor," J. Acoust. Soc. Am. **149**(3), 1454–1465 (2021).

[15] H. Cao, W. Wang, L. Su, H. Ni, P. Gerstoft, Q. Ren, and L. Ma, "Deep transfer learning for underwater direction of arrival using one vector sensor," J. Acoust. Soc. Am. **149**(3), 1699–1711 (2021).

[16] S. Whitaker, A. Barnard, G. D. Anderson, and T. C. Havens, "Recurrent networks for direction-of-arrival identification of an acoustic source in a shallow water channel using a vector sensor," J. Acoust. Soc. Am. **150**(1), 111–119 (2021).

[17] T. S. Brandes, B. Ballard, S. Ramakrishnan, E. Lockhart, B. Marchand, and P. Rabenold, "Environmentally adaptive automated recognition of underwater mines with synthetic aperture sonar imagery," J. Acoust. Soc. Am. **150**(2), 851–863 (2021).

[18] R. A. McCarthy, A. S. Gupta, and M. Kemerling, "Autonomous learning and interpretation of channel multipath scattering using braid manifolds in underwater acoustic communications," J. Acoust. Soc. Am. **150**(2), 906–919 (2021).

[19] Y. Zhang, H. Wang, C. Li, D. Chen, and F. Meriaudeau, "Meta-learning-aided orthogonal frequency division multiplexing for underwater acoustic communications," J. Acoust. Soc. Am. **149**(6), 4596–4606 (2021).

[20] W.-J. Lee and V. Staneva, "Compact representation of temporal processes in echosounder time series via matrix decomposition," J. Acoust. Soc. Am. **148**(6), 3429–3442 (2020).

[21] E. Ozanich, A. Thode, P. Gerstoft, L. A. Freeman, and S. Freeman, "Deep embedded clustering of coral reef bioacoustics," J. Acoust. Soc. Am. **149**(4), 2587–2601 (2021).

[22] E. Cotter, C. Bassett, and A. Lavery, "Classification of broadband target spectra in the mesopelagic using physics-informed machine learning," J. Acoust. Soc. Am. **149**(6), 3889–3901 (2021).

[23] P. Gruden and P. R. White, "Automated extraction of dolphin whistles—a sequential Monte Carlo probability hypothesis density approach," J. Acoust. Soc. Am. **148**(5), 3014–3026 (2020).

[24] P. Gruden, E.-M. Nosal, and E. Oleson, "Tracking time differences of arrivals of multiple sound sources in the presence of clutter and missed detections," J. Acoust. Soc. Am. **149**, 63 (2021).

[25] J. H. Rasmussen and A. Širović, "Automatic detection and classification of baleen whale social calls using convolutional neural networks," J. Acoust. Soc. Am. **149**(5), 3635–3644 (2021).

[26] B. Padovese, F. Frazao, O. S. Kirsebom, and S. Matwin, "Data augmentation for the classification of north atlantic right whales upcalls," J. Acoust. Soc. Am. **149**(4), 2520–2530 (2021).

[27] W. Vickers, B. Milner, D. Risch, and R. Lee, "Robust north atlantic right whale detection using deep learning models for denoising," J. Acoust. Soc. Am. **149**(6), 3797–3812 (2021).

[28] E. Schall, I. Roca, and I. Van Opzeeland, "Acoustic metrics to assess humpback whale song unit structure from the Atlantic sector of the Southern ocean," J. Acoust. Soc. Am. **149**(6), 4649–4658 (2021).

[29] M. A. Roch, S. Lindeneau, G. S. Aurora, K. E. Frasier, J. A. Hildebrand, H. Glotin, and S. Baumann-Pickering, "Using context to train time-domain echolocation click detectors," J. Acoust. Soc. Am. **149**(5), 3301–3310 (2021).

[30] M. Zhong, M. Torterotot, T. A. Branch, K. M. Stafford, J.-Y. Royer, R. Dodhia, and J. L. Ferres, "Detecting, classifying, and counting blue whale calls with siamese neural networks," J. Acoust. Soc. Am. **149**(5), 3086–3094 (2021).

[31] V. Morfi, R. F. Lachlan, and D. Stowell, "Deep perceptual embeddings for unlabelled animal sound events," J. Acoust. Soc. Am. **150**(1), 2–11 (2021).

[32] R. Kuc, "Artificial neural network classification of foliage targets from spectrograms of sequential echoes using a biomimetic audible sonar," J. Acoust. Soc. Am. **148**(5), 3270–3278 (2020).

[33] G. Ciaburro and G. Iannace, "Modelling acoustic metamaterials based on reused buttons using data fitting with neural network," J. Acoust. Soc. Am. **150**(1), 51–63 (2021).

[34] C. Gurbuz, F. Kronowetter, C. Dietz, M. Eser, J. Schmid, and S. Marburg, "Generative adversarial networks for the design of acoustic metamaterials," J. Acoust. Soc. Am. **149**(2), 1162–1174 (2021).

[35] T. Shah, L. Zhuo, P. Lai, A. De La Rosa-Moreno, F. Amirkulova, and P. Gerstoft, "Reinforcement learning applied to metamaterial design," J. Acoust. Soc. Am. **150**(1), 321–338 (2021).

[36] M. Stender, C. Adams, M. Wedler, A. Grebel, and N. Hoffmann, "Explainable machine learning determines effects on the sound absorption coefficient measured in the impedance tube," J. Acoust. Soc. Am. **149**(3), 1932–1945 (2021).

[37] N. Shankar, G. S. Bhat, and I. M. S. Panahi, "Efficient two-microphone speech enhancement using basic recurrent neural network cell for hearing and hearing aids," J. Acoust. Soc. Am. **148**(1), 389–400 (2020).

[38] M. Chinen, J. Skoglund, and A. Hines, "Speech quality estimation with deep lattice networks," J. Acoust. Soc. Am. **149**(6), 3851–3861 (2021).

[39] M. M. Morgan, I. Bhattacharya, R. J. Radke, and J. Braasch, "Classifying the emotional speech content of participants in group meetings using convolutional long short-term memory network," J. Acoust. Soc. Am. **149**(2), 885–894 (2021).

[40] S. Liu, M. Zhang, M. Fang, J. Zhao, K. Hou, and C.-C. Hung, "Speech emotion recognition based on transfer learning from the FaceNet framework," J. Acoust. Soc. Am. **149**(2), 1338–1345 (2021).

[41] M. S. Mahmud, M. Yeasin, and G. M. Bidelman, "Speech categorisation is better described by induced rather than evoked neural activity," J. Acoust. Soc. Am. **149**(3), 1644–1656 (2021).

[42] M. Zhang, X. Pan, Y. Shen, and J. Qiu, "Deep learning-based direction-of-arrival estimation for multiple speech sources using a small scale array," J. Acoust. Soc. Am. **149**(6), 3841–3850 (2021).

[43] R. Riad, J. Karadayi, A.-C. Bachoud-Lévi, and E. Dupoux, "Learning spectro-temporal representations of complex sounds with parameterised neural networks," J. Acoust. Soc. Am. **150**(1), 353–366 (2021).

[44] M. Piotrowska, A. Czyżewski, T. Ciszewski, G. Korvel, A. Kurowski, and B. Kostek, "Evaluation of aspiration problems in L2 english pronunciation employing machine learning," J. Acoust. Soc. Am. **150**(1), 120–132 (2021).

[45] G. Korvel, P. Treigys, and B. Kostek, "Highlighting interlanguage phoneme differences based on similarity matrices and convolutional neural network," J. Acoust. Soc. Am. **149**(1), 508–523 (2021).

[46] N. Ulrich, M. Allassonnière-Tang, F. Pellegrino, and D. Dediu, "Identifying the Russian voiceless non-palatalized fricatives /f/, /s/, and / / from acoustic cues using machine learning," J. Acoust. Soc. Am. **150**(3), 1806–1820 (2021).

[47] N. Tsipas, L. Vrysis, K. Konstantoudakis, and C. Dimoulas, "Semi-supervised audio-driven TV-news speaker diarization using deep neural embeddings," J. Acoust. Soc. Am. **148**(6), 3751–3761 (2020).

[48] C. J. Smalt, G. A. Ciccarelli, A. R. Rodriguez, and W. J. Murphy, "A deep neural-network classifier for photograph-based estimation of hearing protection attenuation and fit," J. Acoust. Soc. Am. **150**(2), 1067–1075 (2021).

[49]M. S. Alavijeh, R. Scott, F. Seviaryn, and R. G. Maev, "Using machine learning to automate ultrasound-based classification of butt-fused joints in medium-density polyethylene gas pipes," J. Acoust. Soc. Am. **150**(1), 561–572 (2021).
[50]A. Abeysinghe, M. Fard, R. Jazar, F. Zambetta, and J. Davy, "Mel frequency cepstral coefficient temporal feature integration for classifying squeak and rattle noise," J. Acoust. Soc. Am. **150**(1), 193–201 (2021).
[51]G. S. Teja, C. S. V. Prasad, B. Venkatesham, and K. S. R. Murty, "Identification of sloshing noises using convolutional neural network," J. Acoust. Soc. Am. **149**(5), 3027–3041 (2021).
[52]Y. Mei, H. Jin, B. Yu, E. Wu, and K. Yang, "Visual geometry group-unet: Deep learning ultrasonic image reconstruction for curved parts," J. Acoust. Soc. Am. **149**(5), 2997–3009 (2021).
[53]N. Liu, H. Chen, K. Songgong, and Y. Li, "Deep learning assisted sound source localisation using two orthogonal first-order differential microphone arrays," J. Acoust. Soc. Am. **149**(2), 1069–1084 (2021).
[54]C. Foy, A. Deleforge, and D. D. Carlo, "Mean absorption estimation from room impulse responses using virtually supervised learning," J. Acoust. Soc. Am. **150**(2), 1286–1299 (2021).
[55]E. Shalev, I. Cohen, and D. Lvov, "Indoors audio classification with structure image method for simulating multi-room acoustics," J. Acoust. Soc. Am. (in press).
[56]J. T. Colonel and J. Reiss, "Reverse engineering of a recording mix with differentiable digital signal processing," J. Acoust. Soc. Am. **150**(1), 608–619 (2021).
[57]H. Pujol, É. Bavu, and A. Garcia, "BeamLearning: An end-to-end deep learning approach for the angular localisation of sound sources using raw multichannel acoustic pressure data," J. Acoust. Soc. Am. **149**(6), 4248–4263 (2021).
[58]D. De Salvio, D. D'Oraziob, and M. Garai, "Unsupervised analysis of background noise sources in active offices," J. Acoust. Soc. Am. **149**(6), 4049–4060 (2021).
[59]D. Ernst Tsokaktsidis, C. Nau, M. Maeder, and S. Marburg, "Using rectified linear unit and swish based artificial neural networks to describe noise transfer in a full vehicle context," J. Acoust. Soc. Am. **150**(3), 2088–2105 (2021).
[60]S. H. Hawley and A. C. Morrison, "Convnets for counting: Object detection of transient phenomena in steelpan drums," J. Acoust. Soc. Am. (to be published).
[61]C. R. Hart, D. K. Wilson, C. L. Pettit, and E. T. Nykaza, "Machine-learning of long-range sound propagation through simulated atmospheric turbulence," J. Acoust. Soc. Am. **149**(6), 4384–4395 (2021).
[62]F. Gontier, V. Lostanlen, M. Lagrange, N. Fortin, C. Lavandier, and J.-F. Petiot, "Polyphonic training set synthesis improves self-supervised urban sound classification," J. Acoust. Soc. Am. **149**(6), 4309–4326 (2021).
[63]H. Chen, Z. Liu, Z. Liu, and P. Zhang, "Long–term scalogram integrated with an iterative data augmentation scheme for acoustic scene classification," J. Acoust. Soc. Am. **149**(6), 4198–4213 (2021).
[64]A. Goudarzi, C. Spehr, and S. Herbold, "Automatic source localisation and spectra generation from sparse beamforming maps," J. Acoust. Soc. Am. **150**(3), 1866–1882 (2021).
[65]J.-R. Gloaguen, D. Ecotière, B. Gauvreau, A. Finez, A. Petit, and C. Lebourdat, "Automatic estimation of the sound emergence of wind turbine noise with non-negative matrix factorisation," J. Acoust. Soc. Am. (to be published).
[66]A. O. Ekpezu, I. Wiafe, F. Katsriku, and W. Yaokumah, "Using deep learning for acoustic event classification: The case of natural disasters," J. Acoust. Soc. Am. **149**(4), 2926–2935 (2021).
[67]V. S. Paul and P. A. Nelson, "Matrix analysis for fast learning of neural networks with application to the classification of acoustic spectra," J. Acoust. Soc. Am. **149**(6), 4119–4133 (2021).
[68]J. Shao, J. Zheng, and B. Zhang, "Deep convolutional neural networks for thyroid tumour grading using ultrasound b-mode images," J. Acoust. Soc. Am. **148**(3), 1529–1535 (2020).