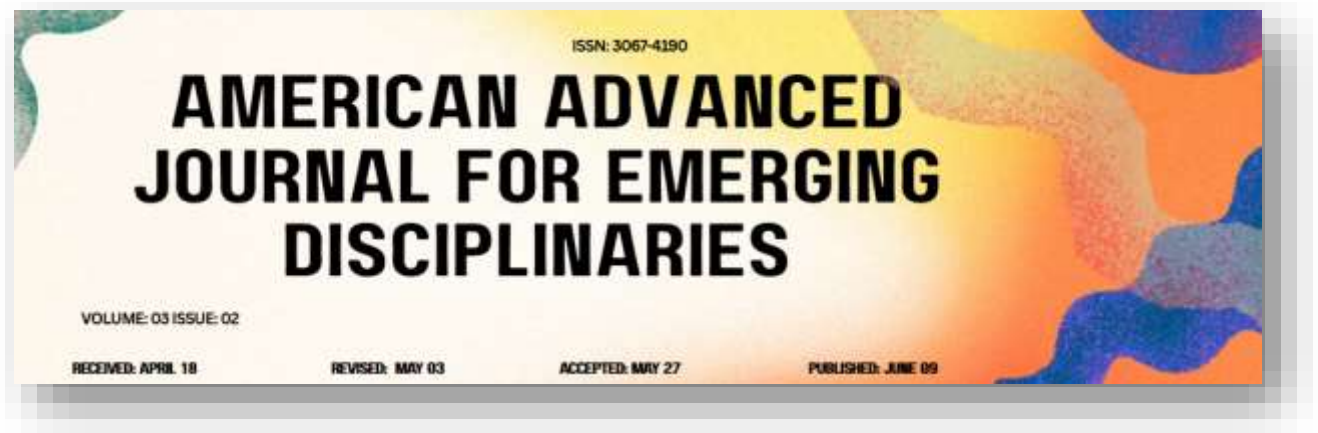




ML-Based Beamforming and Direction Finding

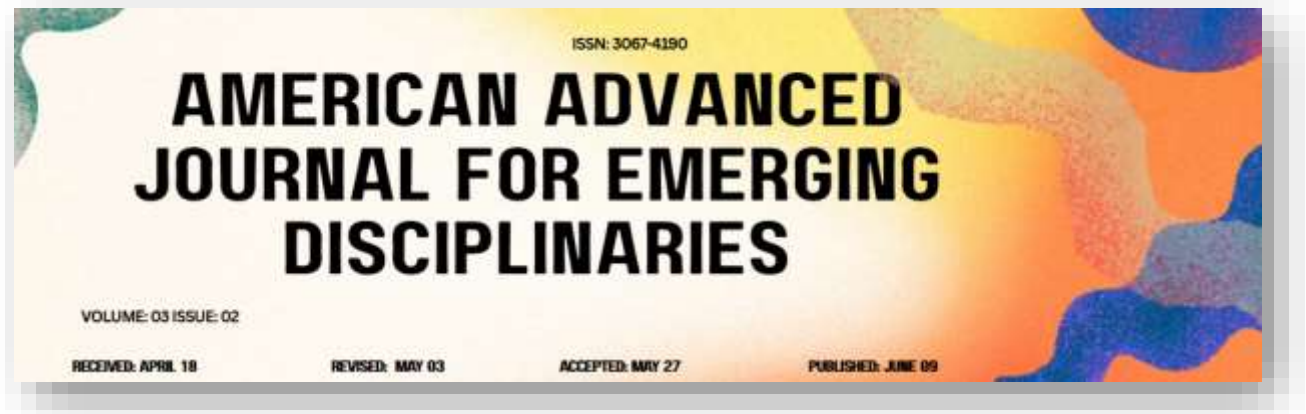
Benjamin Clark
Asst Professor

Abstract

Machine learning techniques for the array signal processing tasks of beamforming and direction finding are reviewed, focusing on recent works in the field. Beamforming techniques for parameter estimation problems are framed as supervised learning tasks with spatial information captured by custom-designed networks. Unsupervised and self-supervised learning paradigms are applied to localized data, tackling tasks with elusive labels such as direction finding in radar and sonar. Evaluation on real-world data brings new problems to the discussion, showcasing the need for domain-oriented solutions. It encompasses networks for convolutional and recurrent networks adapted to spatial signals, transformer architectures for long-range dependencies, and attention mechanisms for depth and efficiency in direction finding. Robustness considerations are addressed, with particular emphasis on adversarial detection and defense.

Machine learning (ML) exploits statistical regularities in data to learn models that are capable of generalizing and performing well under unseen conditions. With years, radar and sonar systems along with their data have grown large. Nonetheless, the underlying abundance of unlabeled data has been overlooked until recently. For some other tasks, data collection is computationally expensive for both labeled and unlabeled data. Such work incorporated, combined, or evolved unsupervised and self-supervised learning paradigms. In supervised settings, the vast majority of approaches model only a small portion of the entire radar/sonar workflow, typically focusing on the parameter estimation problems cast into the beamforming frame. Most studies also utilize 1D temporal represented data. In addition, architectures dedicated to 2D convolutional neural networks have also been adapted to recurrent neural networks on the label correspondences using LSTM units. Modeling across 4D frames brings notable challenges from seeking and capturing spatial long-range dependency with depth and width.

Keywords: Machine Learning For Array Signal Processing, Beamforming Parameter Estimation, Direction Finding Techniques, Supervised Learning For Beamforming, Unsupervised Learning Paradigms, Self-Supervised Learning Methods, Spatial Signal Modeling, Radar And Sonar Systems, Custom Neural Network Architectures, Convolutional Neural Networks For Spatial Data, Recurrent Neural Networks And LSTM, Transformer Architectures



For Long-Range Dependency, Attention Mechanisms In Direction Finding, Real-World Data Evaluation, Domain-Oriented Learning Solutions, Robust Machine Learning Models, Adversarial Detection And Defense, Unlabeled Data Exploitation, High-Dimensional Spatiotemporal Modeling, 4D Signal Representation.

1. Introduction

Machine learning techniques hold great promise for beamforming and direction-finding tasks. Beamforming focuses on the acquisition of spatial information using large aperture linear and 2D arrays to obtain a resolution with a small number of samples. Direction finding aims to recover the embedded steering vectors from a single snapshot of a lower resolution array. Although these two tasks can be intricately linked, they are usually treated in isolation.

In a beamforming setup, multi-channel spatial signals are usually collected across a large aperture, with a potentially significant number of spatial modes but only a few temporal samples. Convolutional neural networks are an intuitive choice for these data, known for their strong performance in similar tasks such as image classification or segmentation. However, these networks are mainly designed to process independent input signals and do not explicitly model the temporal evolution of the spatial signal. To address this limitation and partially incorporate the temporal aspects of the spatial signals, recurrent networks have also been used.

1.1. Overview of the Study and Its Significance

Beamforming and direction-finding methods leverage spatial information and offer high-resolution estimation with low data emanation and processing costs. Significant performance gains can be obtained via data-driven machine learning (ML) techniques, but critical challenges remain. These include the need for large, curated datasets; the evaluation of different ML paradigms; and the exploration of recent advances in architecture design, generalization, robustness, and interpretability. Current research advances these topics.

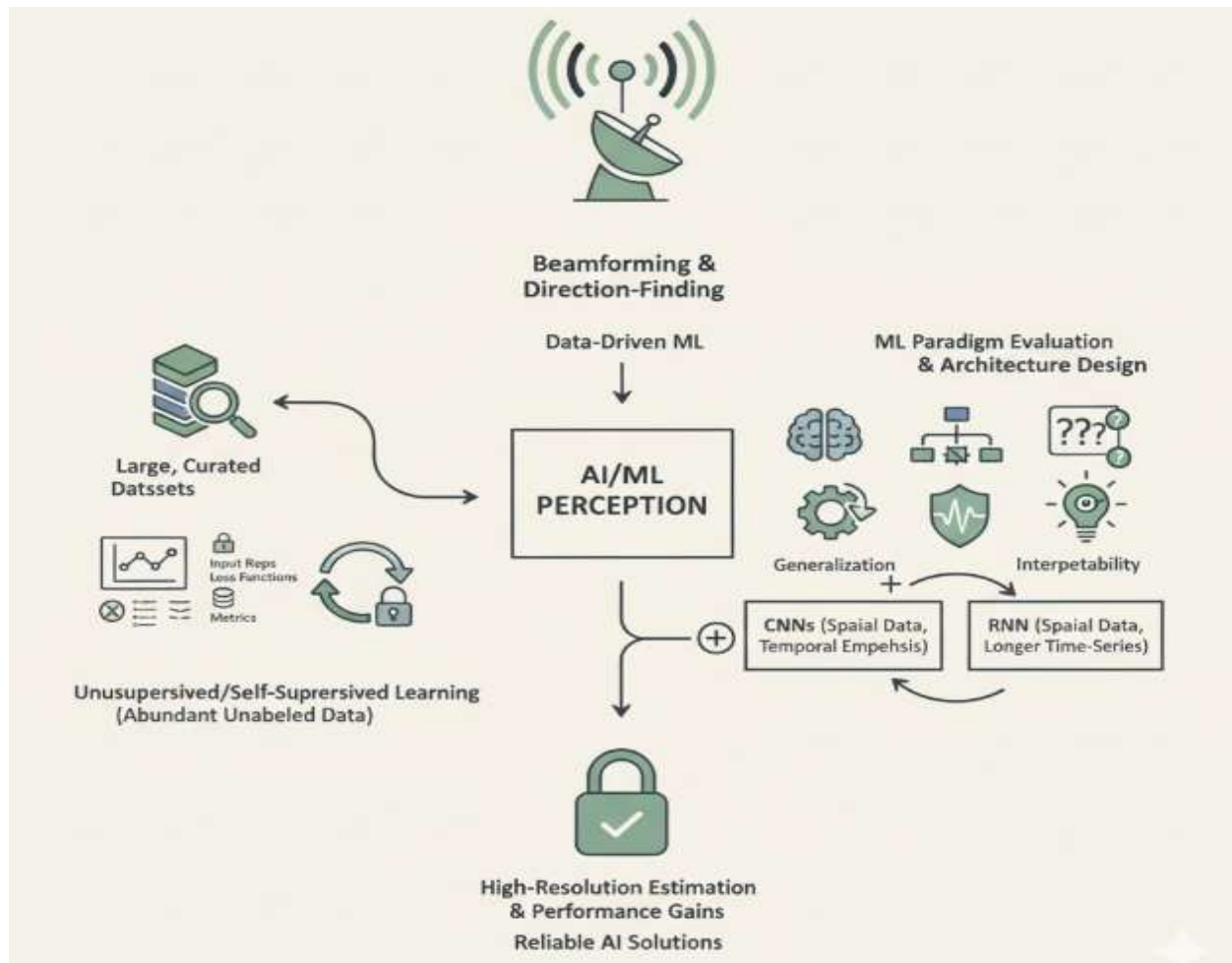
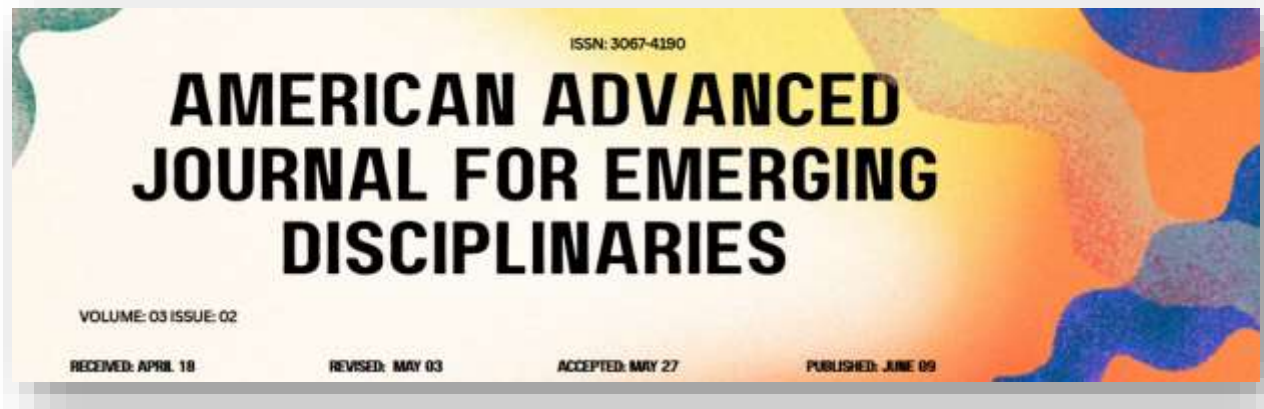


Fig 1: Neural Architecture Evolution for Spatial Perception: Advancing Data-Driven Beamforming and Direction-Finding Systems

Supervised learning for parameter estimation, where the ML model is directly trained to estimate the target parameters, has received considerable attention. Different types of input representations, loss functions, and metrics have been explored, along with the corresponding data requirements, such as volume, labeling effort, and distribution. Moreover, unsupervised or self-supervised paradigms offer opportunities for exploiting abundant unlabeled radar/sonar data

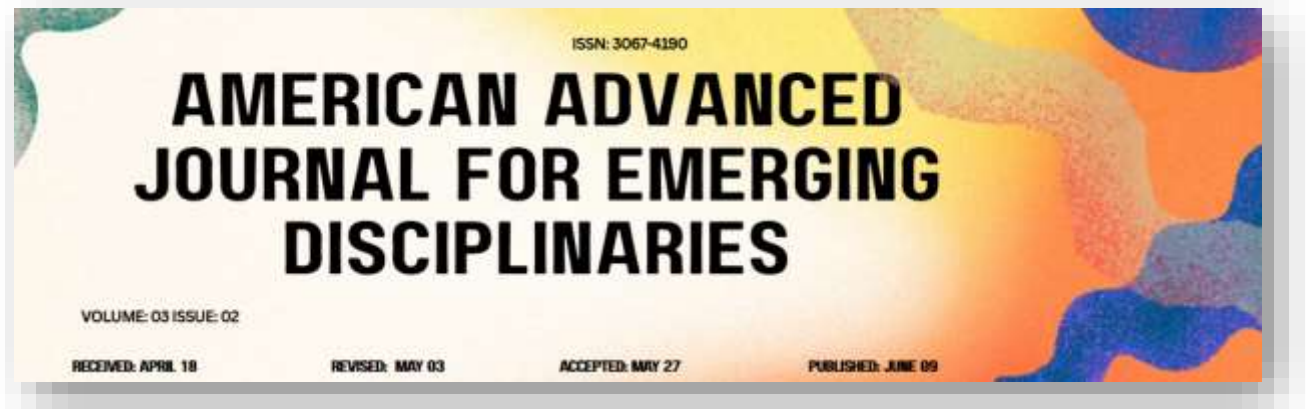


for training perception models. Less emphasis has been placed on beamforming techniques for radar/sonar applications, especially in conjunction with convolutional neural networks (CNNs) and recurrent neural networks (RNNs) applied to the spatial signals. Both CNNs and RNNs analyze the spatial data but, conceptually, in different ways. CNNs model the temporal dimension more easily than RNNs, whereas RNNs are better suited for longer time-series foreshortening and/or extending sampling intervals between time frames. Pursuing the latter function with CNNs typically requires larger datasets and sacrifices some generalization ability.

2. Fundamentals of Beamforming and Direction Finding

Beamforming is an array signal processing method that aims to use received sensor signals for parameter estimation related to source activity, such as direction of arrival (DoA) or time delay of arrival (TDoA). It realizes spatial filtering by creating a bank of directed filters along selected parameter hypersurfaces, which enhances signal contribution from targeted clusters while suppressing undesired interference and noise components. The spatial frequencies of signals in the received data are defined via the steering matrix, each column being the steering vector for the corresponding parameter point. The source activity can then be represented by a low-rank approximation of the data covariance matrix using its singular value decomposition, which provides a data-adaptive parameterization of the physical noise-free array manifold. Conventional beamformers employ a straightforward reduce-extract-regularize strategy based on the estimated steering matrix. From the DoA estimation perspective, the minimum variance distortionless response (MVDR) beamformer is a standard choice. The known covariance structure along with the MVDR beamform is then used to define the loss function for supervised learning.

Traditional beamforming methods are through-the-radar/sonar images with limited resolution, in particular when using small sensor arrays or under low-sensor-density conditions. The spatio-temporal Fisher information matrix, which is a generalization of the conventional Fisher information matrix to vector-valued spatio-temporal signals, is often used to define resolution limits. Broadly speaking, blind source separation, DoA estimation, TDoA estimation, and 3-D source localization can all be performed on time-delayed processed radar or sonar data with. The joint exploitation of both spatial and temporal dimensions in a single beamforming method has been recently considered. Given a temporal signal X sampled from a narrow-bandpoint source with unknown DoA and with additional uncorrelated wideband noise, a



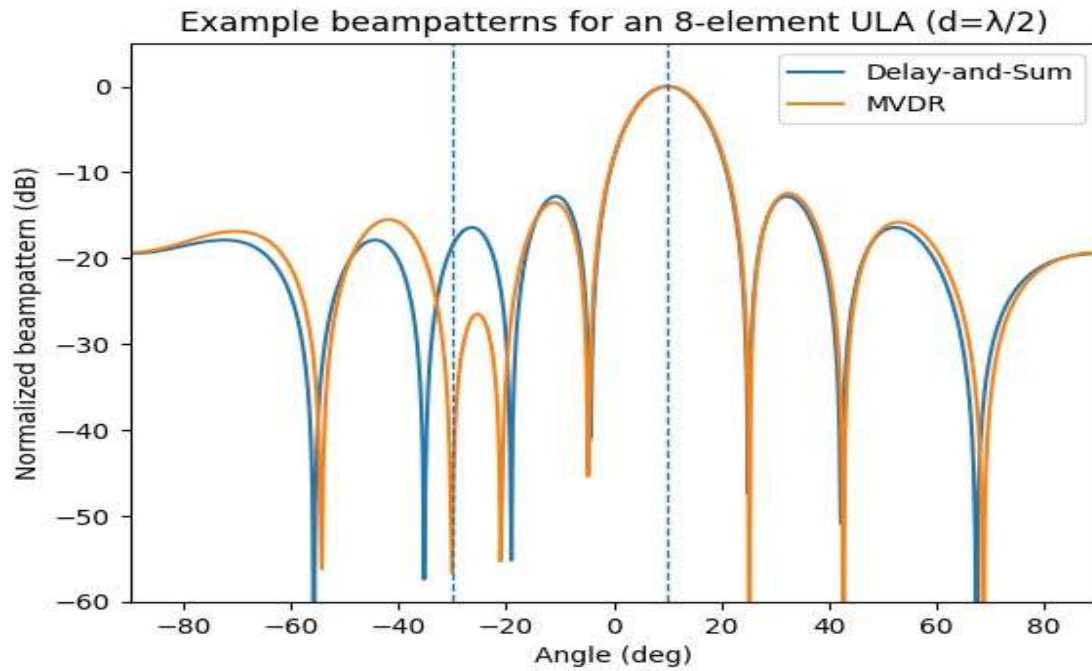
convolutional neural network (CNN) is trained to emulate the MVDR beamformer. The model takes the raw temporal data as input and directly reconstructs the spatial image of the source.

2.1. Array Signal Processing Essentials

The array signal processing framework is first recapped. Definitions of steering vectors, array response manifolds, and resolution conditions are then provided. Typical targets for parameter estimation, considered when using data-driven or model-free methods, are also discussed.

An array is comprised of sensors arranged in a geometrical configuration (line, circle, etc.). The signals received by an array exposed to a monochromatic plane wave traveling from a specific direction span a steering vector. In the case of narrowband sources, the vector has a length equal to the number of sensors elements in the array. When a set of concurrent signals share the same frequency this array response vector is in turn part of a matrix: the array response manifold. These mathematical concepts summarize important information when training and testing methods aimed at estimating the direction of arrival of a source impinging on a sensor array. For a given array configuration and a monochromatic plane wave arriving from a certain direction, the output signal-to-noise ratio of a certain beamforming method behaves as an indicator function defined by the steering vector.

Standard techniques relied upon to estimate the direction of a source are the Minimum Variance Distortionless Response (MVDR) beamformer and the delay-and-sum method. These approaches can yield unsatisfactory results when external noise sources are not white, when the sources are not sufficiently separated in angle, or when the steering vectors are not known a priori. These issues are closely related to the quality of the training and test datasets. The completeness of the datasets for supervised learning is instrumental to guaranteeing performance when applied to new situations.



Equation 1: Signal model and steering vector (array manifold)

1.1 Narrowband snapshot model

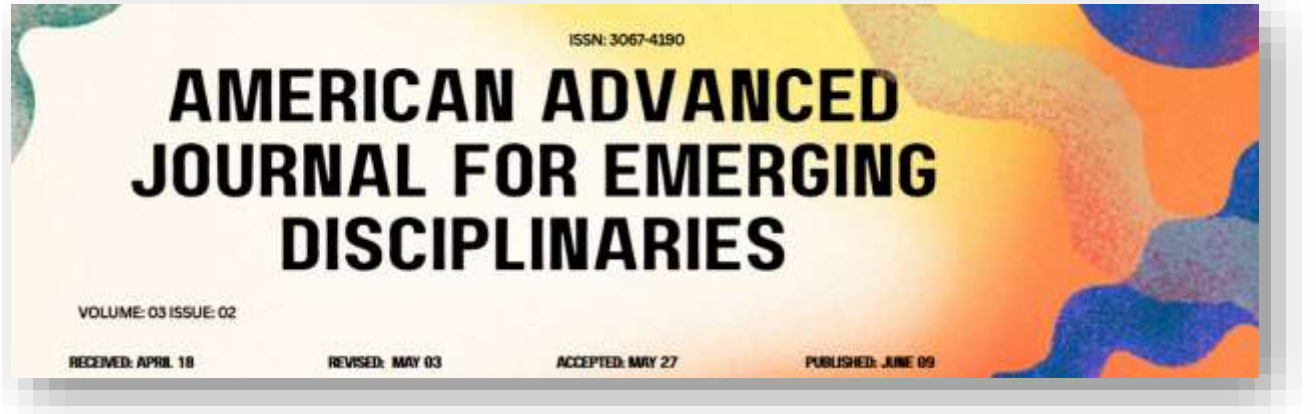
Assume M sensors, and we observe a **single time snapshot** (or a time index t):

$$\mathbf{x}(t) \in \mathbb{C}^{M \times 1}$$

For a single narrowband source from direction θ :

$$\mathbf{x}(t) = \mathbf{a}(\theta)s(t) + \mathbf{n}(t)$$

- $\mathbf{a}(\theta)$: steering vector ($M \times 1$)
- $s(t)$: complex source amplitude



- $\mathbf{n}(t)$: noise (+ interference)

For K simultaneous sources:

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t)$$

where

$$\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)].$$

1.2 ULA steering vector (common closed form)

For a **Uniform Linear Array (ULA)** with sensor spacing d , wavelength λ , and broadside reference:

1. Path difference to sensor m is proportional to $md\sin\theta$
2. Phase shift is $2\pi \frac{md\sin\theta}{\lambda}$

Thus, for sensors $m = 0, \dots, M - 1$:

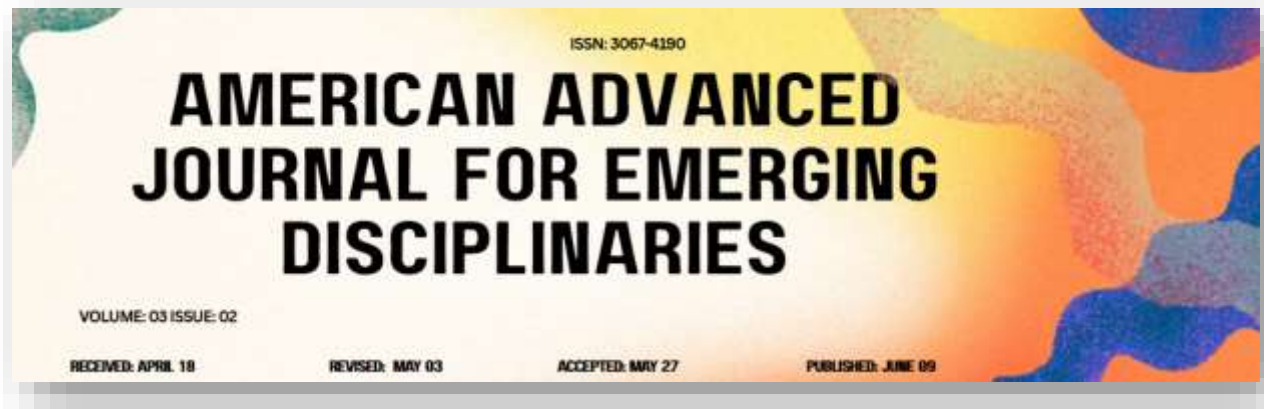
$$a_m(\theta) = e^{j2\pi \frac{d}{\lambda} m \sin\theta}$$

So

$$\mathbf{a}(\theta) = \begin{bmatrix} 1 \\ e^{j2\pi \frac{d}{\lambda} \sin\theta} \\ \vdots \\ e^{j2\pi \frac{d}{\lambda} (M-1) \sin\theta} \end{bmatrix}$$

2.2. Traditional Beamforming Techniques

Classic approaches to beamforming and direction finding employ a distinct mathematical structure and are based on particular design criteria. While minimum variance distortionless response and delay-and-sum beamformers are often considered, other techniques also exist.



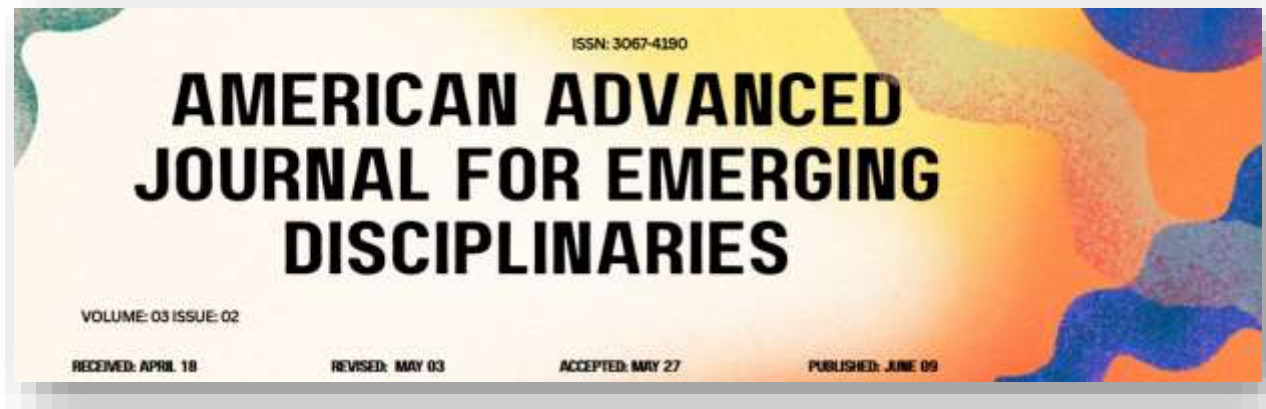
Comparison of these methods reveals specific drawbacks that may be addressed by employing an alternative beamforming perspective, leading to a more general design paradigm.

The minimum variance distortionless response algorithm minimises the expected mean square noise output power of the beamformer while guaranteeing unit gain on the desired signal, thus having the potential to provide superior performance in noisy environments. However, when the incoming signal is not located precisely on the directional axis of the array, the remaining non-zero response of the beamformer will be subject to distortion. In contrast, the delay-and-sum beamformer has a simple and straightforward structure that allows real-time implementation and operation below the Nyquist sampling rate. It does not distort the incoming signal; however, in a noise-dominated environment it provides inadequate noise suppression and is prone to interference corruption.

Method	Interference angle (deg)	Response at interference (dB, normalized)	Approx. -3 dB beamwidth (deg)
Delay-and-Sum	-30.0	-18.602257781092714	13.25
MVDR	-30.0	-56.653722473279934	13.5

3. Machine Learning Paradigms for Array Processing

Category-based classification of machine learning models can also be made in terms of presence or absence of explicit supervision during the training process. In supervised learning, a given task is associated with the ground truth, which is a categorical label for classification tasks or a target prediction for regression problems. In most cases, a sufficient amount of training data is not available for large, complex perceptron-type or other deep structures to generalize. Hence, unsupervised learning employs only input data for training, in a totally unlabeled manner, with the help of some auto-encoding networks and cross-domain-fusion-based generative networks—taking random noise from various data domains as input and generating raw input data through proper supervision. The trained model can generate data in various domains. In a more intelligent version, self-supervised learning enhances training using supervision signals generated from within the data, with labeled training data obtained in a semi-supervised approach. USSL is



useful for producing self-supervised representation from a multi-task mixture setup and bridging learning through unlabeled, extremely low-cost data with a small amount of supervised data.

In the specific context of array engine training, most methods based on supervised learning focus directly on the beamforming output, which indicates where the target is located or how the target manifests in the received data. Although the raw spatial-temporal training signal may help train a perceptor network in an unsupervised manner, the generation of raw spatial-temporal data for radar using other problems is not straightforward. Thus, direct unsupervised learning for the spatial-temporal array data still requires additional efforts until a fully automatic data-labeling method emerges. It is currently possible to apply semi-supervised learning to develop a weakly supervised mechanism for radar's very scarce visual data, as the visual domain provides an abundant amount of unlabeled data. For oceans, far infrared, and other difficult environments, large-scale radar/sonar data still remain unexplored for unsupervised learning. Hence, these raw spatial-temporal data with very few labeled data are leveraged in a self-supervised manner to construct suitable model representation.

3.1. Supervised Learning for Parameter Estimation

Parameter estimation is framed as a supervised learning problem, with the steering vector of an input space spanned by the manifold. A fully supervised approach requires labeled data when training the neural network estimator. Typical inputs to the network are the spatial domain samples arranged in a 2D representation with one spatial dimension for time and the other for azimuth while the label generated is of a heatmap representation with one or more peaks. Which peak is to be considered as the label will depend on the loss function being used. L1 or L2 loss can be used when considering a direct estimation of the steering vector whereas more appropriate loss functions will be needed for the heatmap representation. Typical evaluation metrics are the mean prediction localization error for a single peak or the mean localization accuracy for K peaks found in the top locations of the spatial domain samples being evaluated. These two

metrics are complementary.

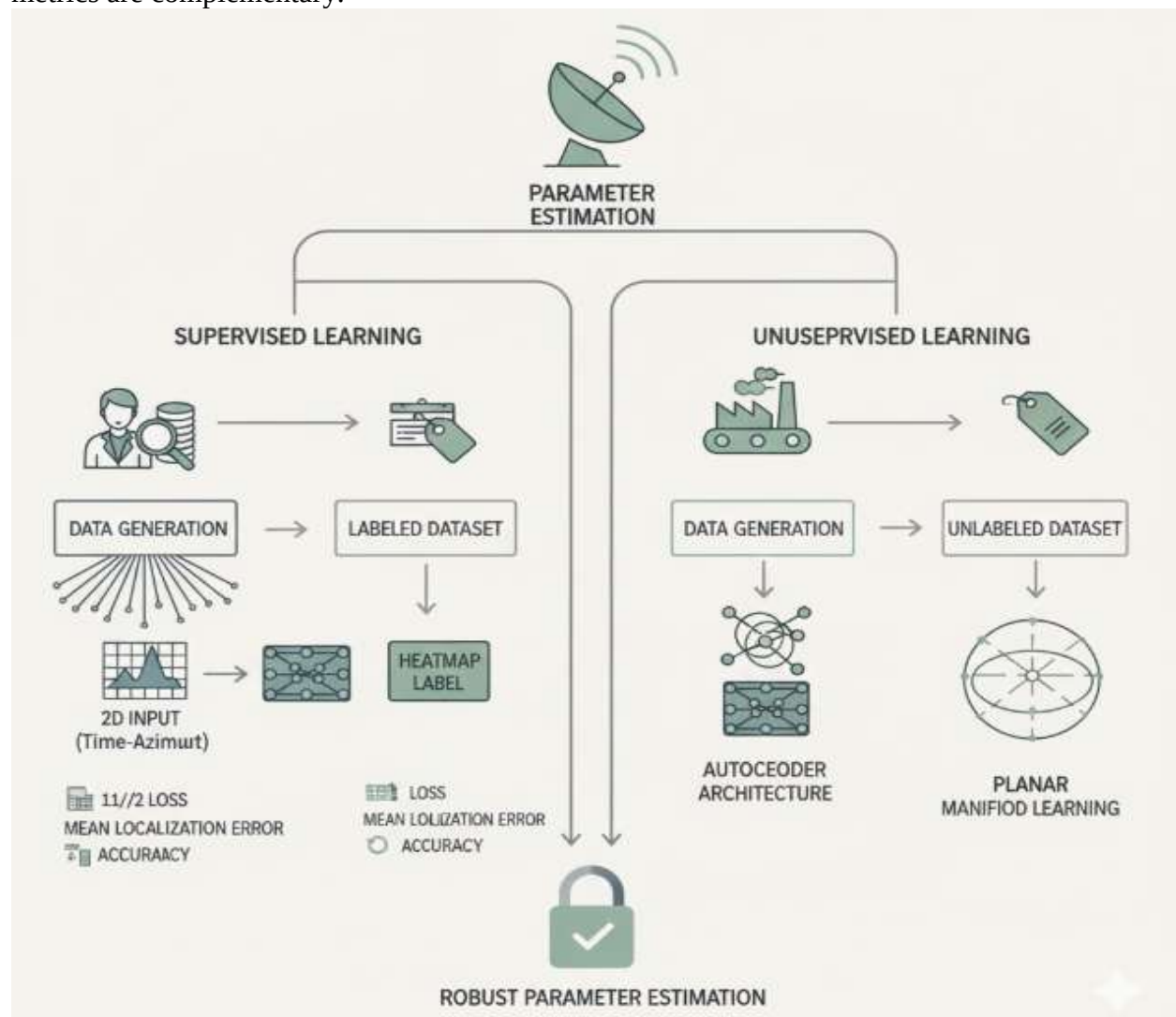
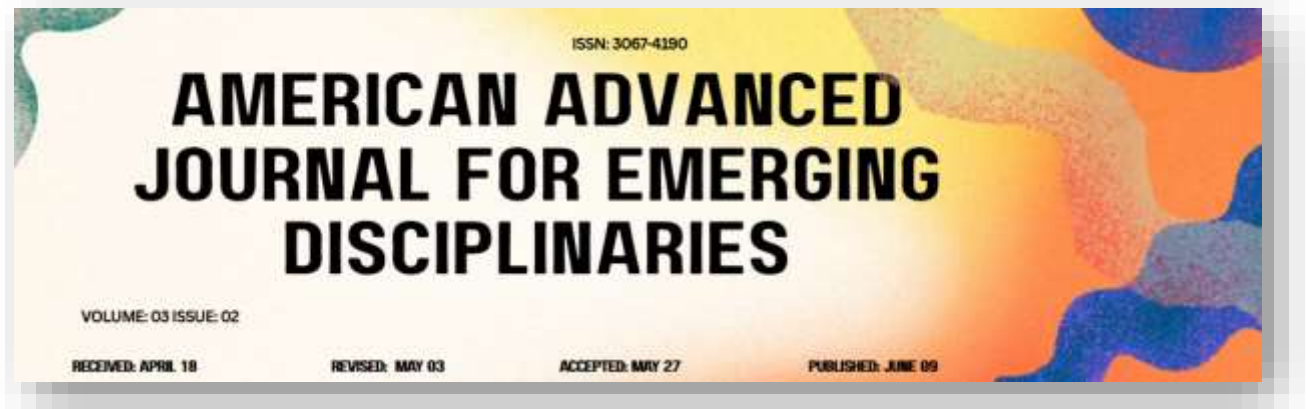


Fig 2: Manifold-Aware Learning Paradigms: From Supervised Heatmap Estimation to Unsupervised Autoencoders in Radar and Sonar Sensing

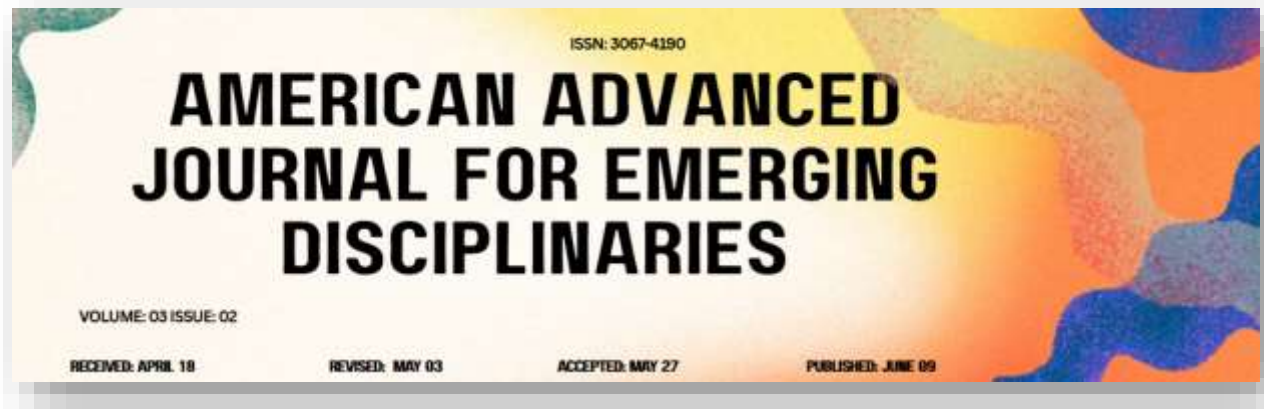


Designing an appropriate data generation pipeline capable of generating data covering the whole span of the considered parameter space with sufficient details is key for obtaining good network performance. However, a fully supervised approach requires the generation of a large amount of labeled data, which can be expensive in many real-world scenarios especially in a radar or sonar sensing context. A completely unsupervised approach relies on a data generation pipeline eliminating the labeling step during the data generation process making possible to generate a very large data sample. The challenge is then to design a proper autoencoder architecture capable of learning the underlying spatial parameterization without resorting to labeling the data. A representation that takes advantages of the planar nature of the bistatic radar or sonar steering vector can be exploited to design an actual autoencoder.

3.2. Unsupervised and Self-Supervised Approaches

Unsupervised and Self-Supervised Approaches. A fundamental assumption of supervised learning is that labeled data are available for both training and testing the model. Unsupervised learning of the data distribution can address the label scarcity issue, allowing the algorithm to learn from an unlabeled dataset instead of using labeled examples for training or fine-tuning. An objective function measuring the similarity of the model-reconstructed output with the original input signal can represent the correlation structure of signals, specifically the joint probability distribution of the input samples. In array signal processing, however, the desired correlation structure is weighted by the array manifold, whose information is not included in unlabeled data. In arrays processing, unsupervised learning can, nevertheless, capture disorderly spatial signals (e.g., sonar and radar without strong targets or noise, or calamity-like radar signals) for spatial similarity in parameter estimation. The structure can also assist the supervision of limited labeled sounds (e.g., bird recognition) or events (e.g., fusion source separation).

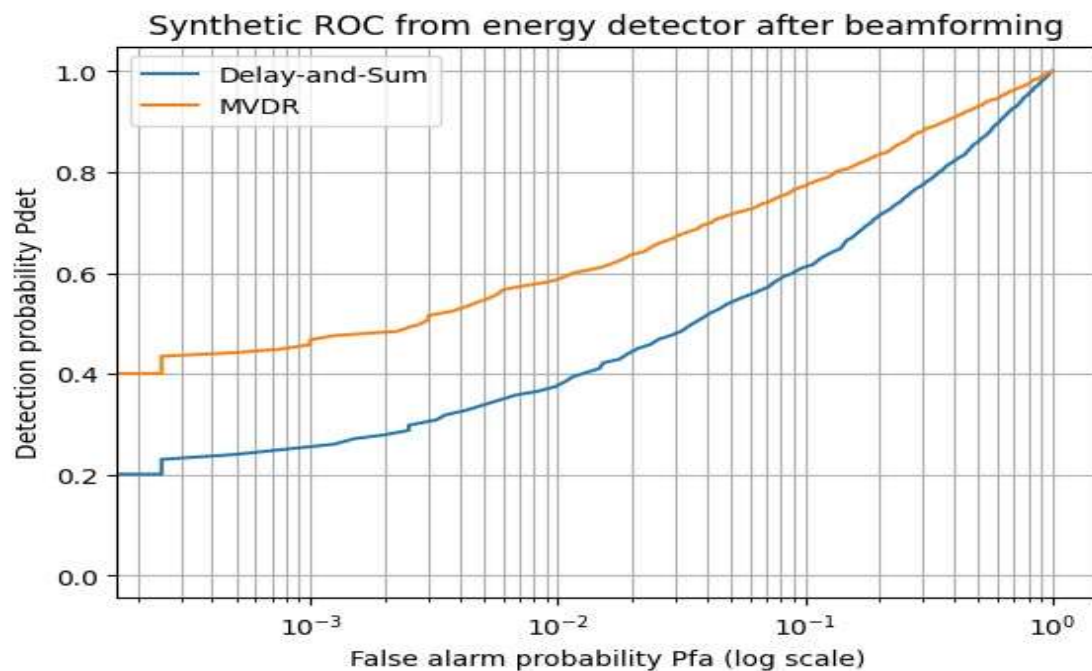
Self-supervised learning approaches aim to learn a feature representation for a downstream task by creating a surrogate label from the input data, thus bridging the gap between unsupervised and supervised learning. Temporal objectives can exploit the structure of temporal signals, while the data augmentation property of label perturbation is useful for minimizing overfitting and ensuring domain-generalization capabilities in classification tasks. Two main types of learning objectives can be derived from the probability distribution of surface-arranged sensors: contrastive and predictive tasks. Contrastive tasks utilize the pairwise relationship of the spatial data recorded at the same time, while predictive tasks explore the temporal property of spatial data to make predictions on the future position of the signals.



4. Data, Simulation, and Evaluation

The dataset of supervised approaches comprises pairs of input-output data samples. For supervised parameter estimation methods, soft- and hard-label data was extracted from Monte Carlo simulations. In particular, labeled data were constructed by collecting 5D matrices of size (f, θ, ϕ, g, n) that correspond to the signals received from a point-source target (i.e., steering vectors) centered in the angular direction (θ, ϕ) and at frequency f of the received radar pulse train. A data generation pipeline simulating radar/sonar scenarios was specifically implemented to provide realistic conditions by including the non-idealities of real data. These conditions are crucial to guaranteeing fair comparison with the conventional algorithms when both the traditional methods and the deep-learning models are evaluated on the same test sets, rolled out approximately from the same Monte Carlo simulation. A move toward using the present methods on spatial data requires caution in terms of training realism.

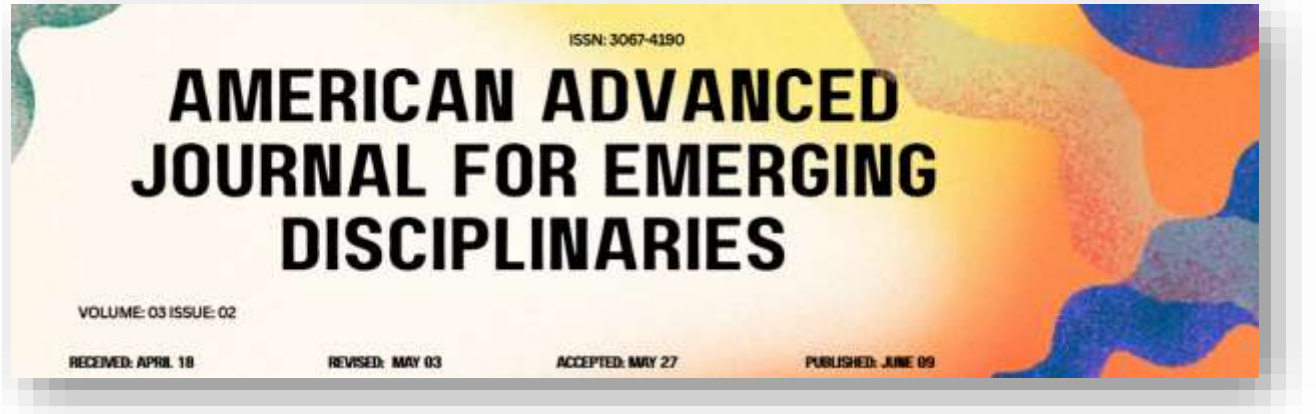
For beamforming performance assessment, detection probability (P_{det}), false alarm probability (P_{fa}), mean squared error (MSE), and structural similarity (SSIM) are computed. P_{det} and P_{fa} are derived from the benchmarked detection routines executed on the output of both conventional methods and the trained deep networks. Both probabilities proceed over a set of N_{test} samples, where a detection threshold θ is set up considering that the probabilities are complementary. MSE is calculated on the ambiguous maps used for beamforming. Finally, SSIM is computed over the soft-mask representation of the results, with scale factor S set to 0.5, complementing the ambiguous behavior considering it with respect to the original panchromatic image. The multi-fold implemented pipeline is thus capable of rigorously training deep networks from scratch on data that would be otherwise sparsely labeled.



4.1. Datasets and Benchmarking Protocols

For supervised learning approaches, training datasets must reflect probable operational conditions while modeling the intended predictions. Whether designed for deep learning or not, the data generation pipelines and results contain arbitrary and, where possible, realistic sources of mismatch. In all cases, separate validation and testing datasets assess performance relative to conditions that were not simulated in the training data.

Four radar/sonar benchmark datasets for supervised learning are introduced, all generated with the open-source simulator MIMO-Sim. The first two datasets feature moving vessels around and within a port. The third consists of a group of ships, several of which are moving. The classification task for the sonar dataset contains real data external to the simulator. A common set of metrics is implemented, ensuring consistent evaluation quality across different works. The complete datasets are made publicly available for training, validation, and testing.



Equation 2: Covariance matrix and why it matters

Define the spatial covariance:

$$\mathbf{R} = \mathbb{E}[\mathbf{x}\mathbf{x}^H]$$

Using $\mathbf{x} = \mathbf{a}s + \mathbf{n}$ and assuming s independent of $\mathbf{n}$, zero-mean, and $\mathbb{E}[|s|^2] = \sigma_s^2$:

$$\mathbf{R} = \mathbb{E}[(\mathbf{a}s + \mathbf{n})(\mathbf{a}s + \mathbf{n})^H]$$

Expand:

$$\mathbf{R} = \mathbb{E}[\mathbf{a}s s^* \mathbf{a}^H] + \mathbb{E}[\mathbf{a}s \mathbf{n}^H] + \mathbb{E}[\mathbf{n} s^* \mathbf{a}^H] + \mathbb{E}[\mathbf{n} \mathbf{n}^H]$$

Cross-terms vanish by independence and zero-mean assumptions:

$$\mathbb{E}[\mathbf{a}s \mathbf{n}^H] = \mathbf{a} \mathbb{E}[s] \mathbb{E}[\mathbf{n}^H] = 0$$

So:

$$\boxed{\mathbf{R} = \sigma_s^2 \mathbf{a} \mathbf{a}^H + \mathbf{R}_n}$$

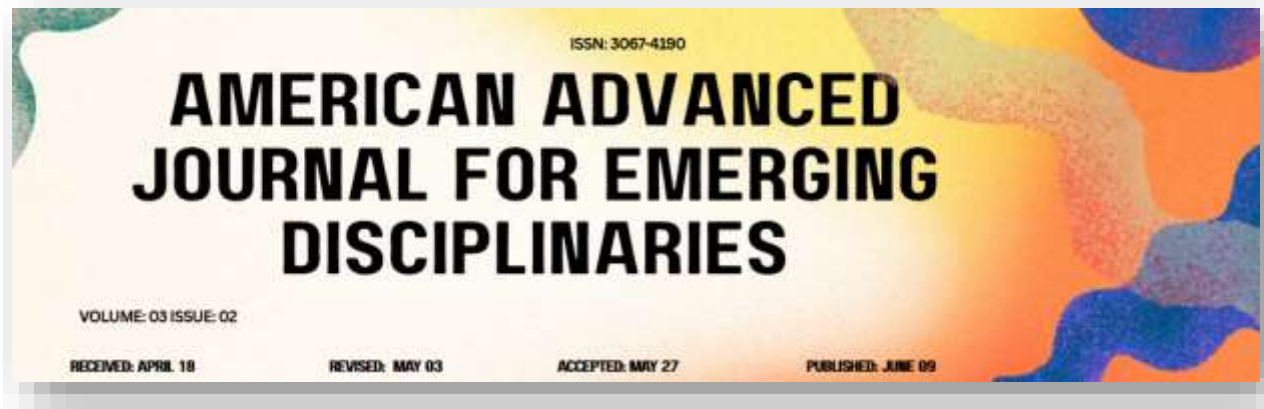
If noise is spatially white: $\mathbf{R}_n = \sigma_n^2 \mathbf{I}$.

In practice with N snapshots:

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{t=1}^N \mathbf{x}(t) \mathbf{x}(t)^H$$

5. Deep Learning Architectures for Beamforming

Two key themes emerge in recent neural network applications to beamforming—convolution or recurrent processing of spatial signals. Deep neural networks are often interpretable as multiscale convolutional networks and have been applied directly to the spatial channels (magnitude and phase). These networks expand upon conventional techniques that model the spatial aspect of the data with a convolution while treating the temporal aspect with a recurrent network. An alternative approach uses temporal analysis in the choice of kernel, at the cost of increased model size and potential inefficiency due to feature redundancy when processing



each time point separately. The temporal representation is more efficient for learning from an environment with a strong temporal dependence, spatially-coherent sources, or lower-SNR signals, while the spatial kernel is more powerful for learning from unlabeled data sets.

These considerations are largely orthogonal to those of the previous section; greater temporal extent, larger model size, and unsupervised learning would increase the capacity of the model to learn from spatially-distributed temporal signals such as sonar backscatter data. Attention mechanisms have recently gained significant traction in the deep learning community, with impressive results in many application areas. Transformers and attention-based architectures enable sophisticated incorporations of long-range temporal and spatial dependencies, generally offering reduced training time and resource requirements relative to recurrent neural networks. A key advantage in the present context is amenability to large arrays with long-range spatial dependencies, such as communication with motherships and group arrivals with multiple-delay superposition.

5.1. Convolutional and Recurrent Networks for Spatial Signals

Convolutional and recurrent networks have been developed for beamforming and direction-finding with multichannel signals observed over time. Convolutional layers leverage the spatial structure of the data, acting on either the entire time-frequency representation or on its individual channels (time series). The former enables direct estimation of the beamforming output, while the latter also supports a second network stage that processes the outputs of standard beamformers. Recurrent architectures exploit long short-term memory units to capture temporal order and correlations along the time axis. When adequate training data are available, they have also been shown to outperform standard beamformers even when trained directly on their inputs for time-frequency representations. However, the resulting large networks typically require many samples per training example, especially compared with spatially-structured architectures.

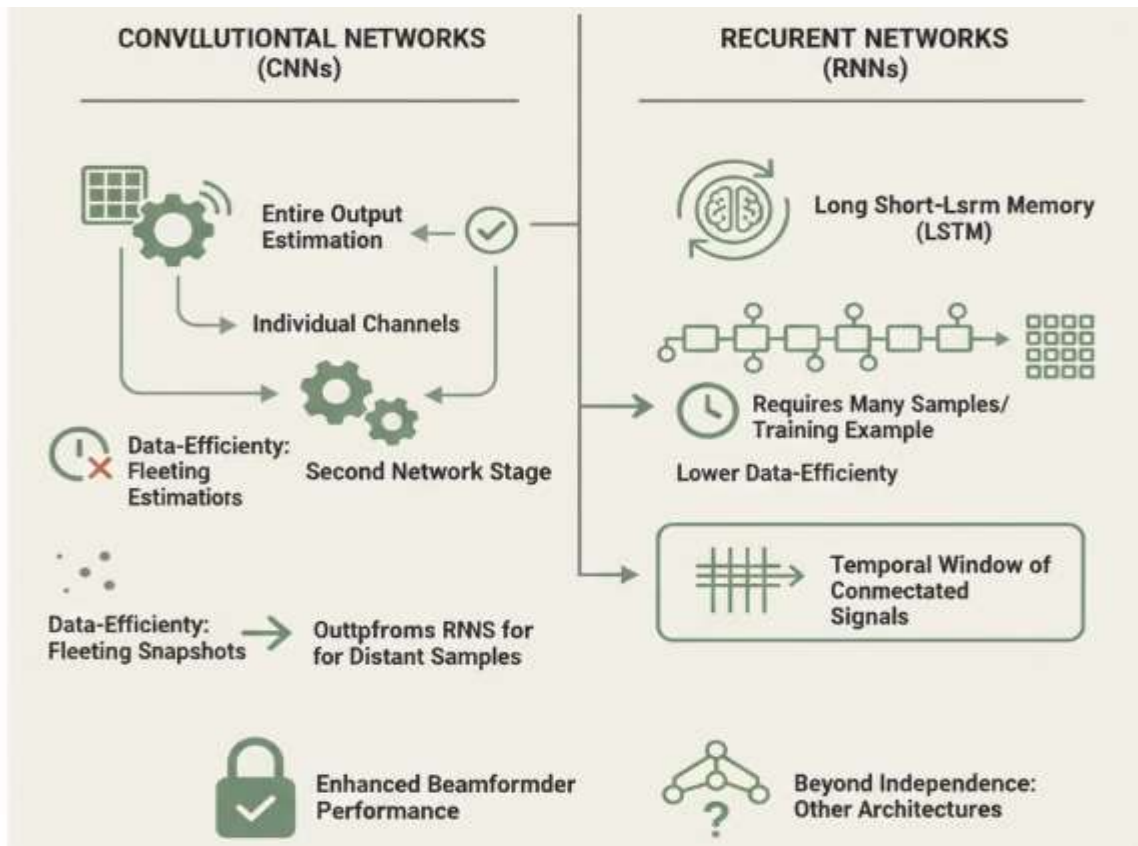
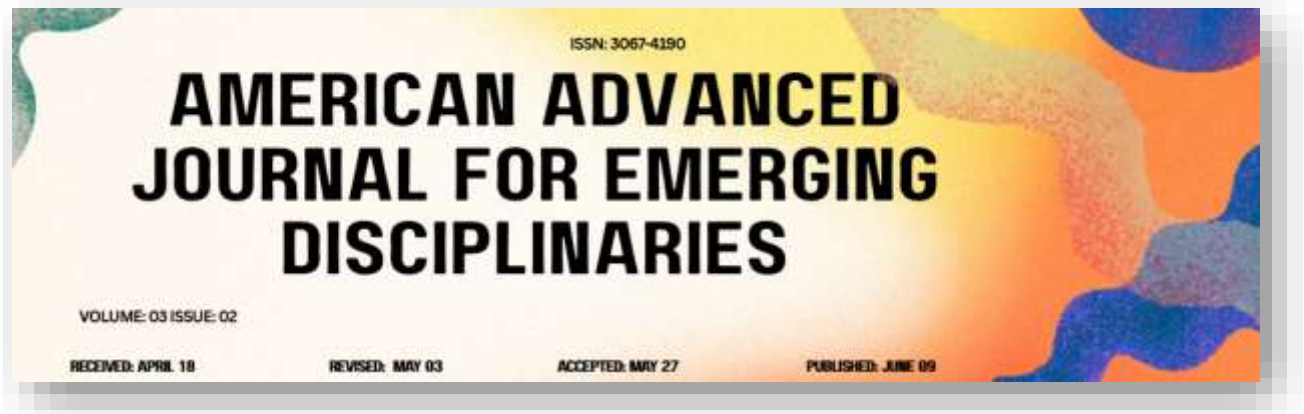


Fig 3: Spatio-Temporal Trade-offs in Deep Learning for Beamforming: Evaluating CNN and RNN Architectures for Data-Efficient

Indeed, from a data-efficiency perspective, direct estimation based on fleeting snapshots is significantly more favorable than a temporally-ordered sequence of independent estimations. Convolutional neural networks thus clearly outperform recurrent architectures when training samples are distant from one another. One option for addressing this limitation is to use recurrent layers operating on a temporal window of concatenated channel signals, which ensures that a reduced number of estimations can jointly inform the result from the final time step. Nevertheless, while exploiting temporal correlations, this approach fails to take advantage of temporal coherence, and its critical temporal resolution essentially remains governed by the



distance between individual snapshot samples. When working below the standard model, other architectures allow directly relaxing the independence assumption in temporal series.

5.2. Attention Mechanisms and Transformer-Based Models

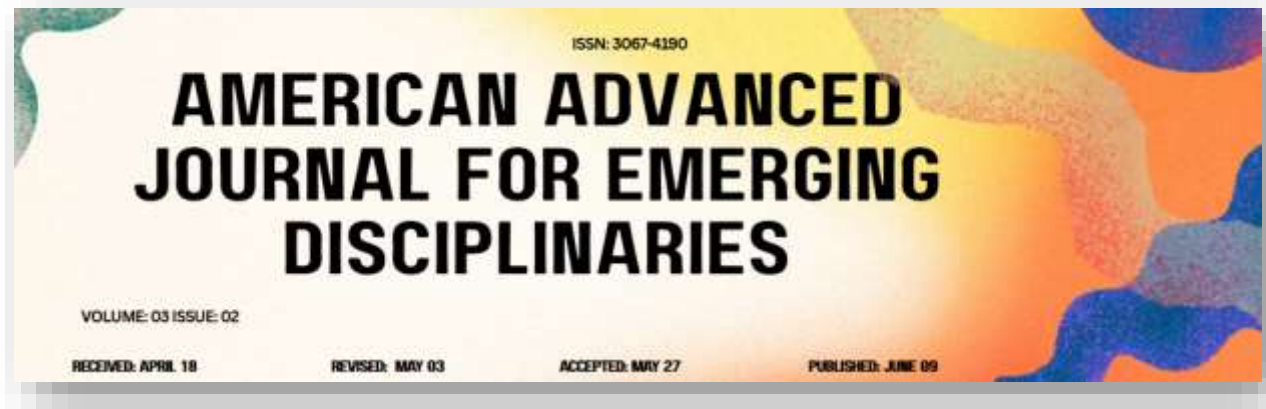
Attention mechanisms capture interrelations between input elements, enabling the selective weighting of spatially-distributed signals for deeper modeling. These modules improve data efficiency and support long-range reasoning but typically require extensive labeled datasets. Transformers, driven by self-attention and circumventing recurrence, have attained remarkable performance across many machine-learning domains. Their parallelizability, scalability, and representational power make them appealing for signal processing. Attention-based paradigms have thus emerged for direction-finding networks.

Self-attention models have been leveraged in direction finding and other high-dimensional supervised tasks. The global modeling of self-attention, while highly expressive, traditionally scales cubically with input size. Long-Range Attention alleviates this limitation by relating signals only to nearby inputs, with matching capabilities for far-flung signals. While attention modules augment many architectures, their success in end-to-end networks remains largely unexplored. In particular, the integration of self-attention within transformer networks has yet to be examined for direction-finding applications. A separate branch of literature investigates multi-head Attention in the context of amodal generation tasks, where input signals are insufficient for complete recovery. These pioneering works, along with extensive research in the natural language domain, inspire the exploration of attention mechanisms and transformer models for direction finding.

Method	Pdet at $P_{fa} \approx 1e-2$
Delay-and-Sum	0.37425
MVDR	0.58675
CNN-emulated MVDR (illustrative)	0.55675

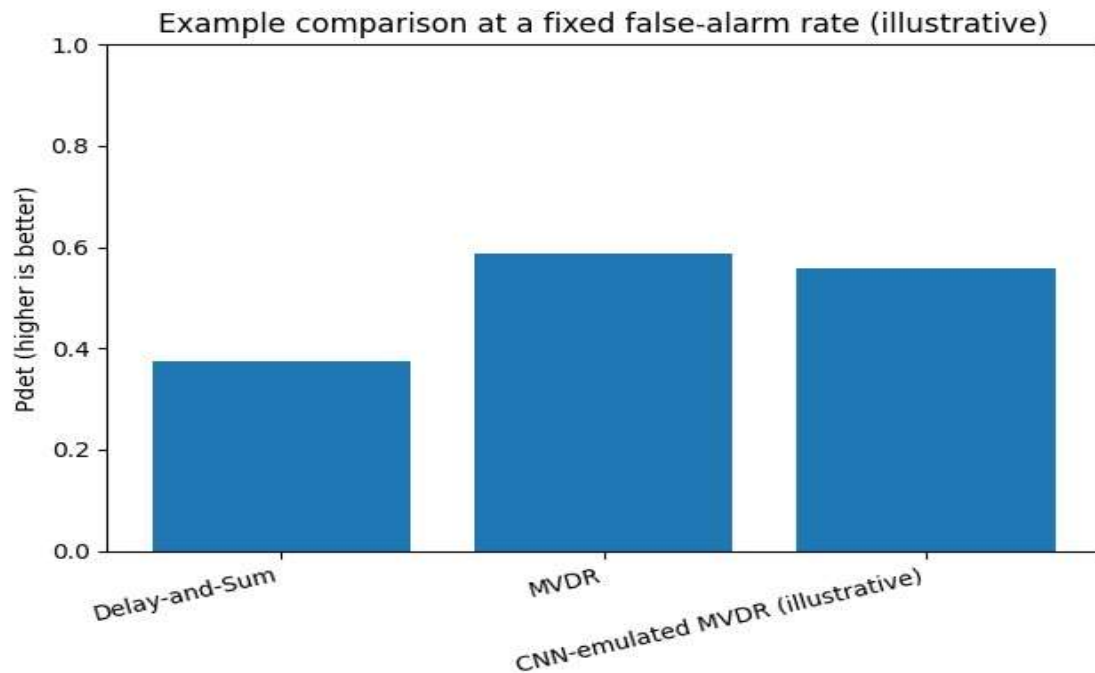
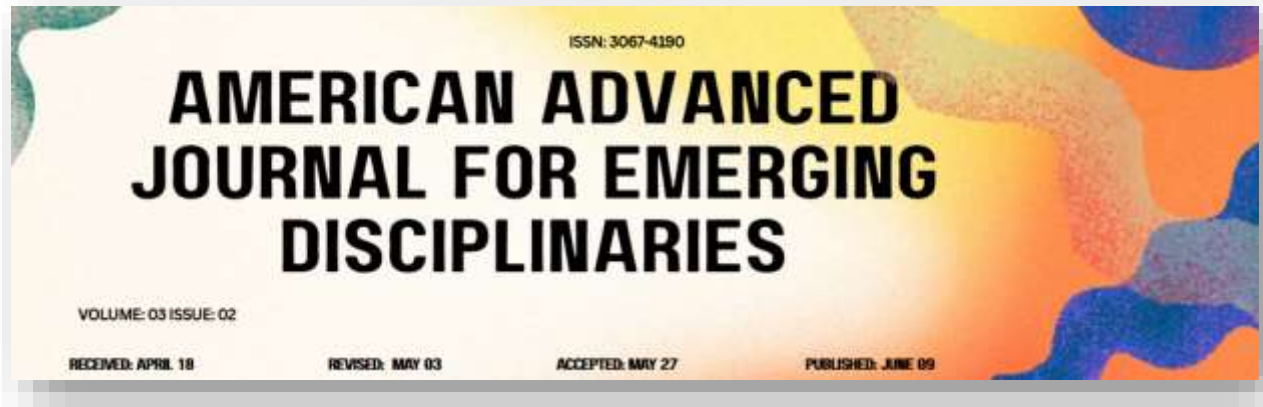
6. Robustness, Generalization, and Adversarial Considerations

Mismatch between training and testing conditions negatively influences the robustness and generalization capabilities of beamformers. Mismatches can be induced, for instance,



through the mismatching of spatial correlation between sources and noise, the addition of unknown spatially correlated noise, or the randomization of the steering vectors. Furthermore, the presence of noise or varying noise conditions affects the number of bits used for quantizing the phase and amplitude of the uncertain steering vector. Such considerations ensure that neural networks maintain their robustness away from the training conditions, and various techniques, such as dropout layers, were incorporated in the networks during training to increase generalization. Additionally, adversarial attacks against neural network-based beamformers were employed to achieve the best possible performance. Both white-box and black-box adversarial examples were taken into account during evaluation, and defensive distillation techniques were applied to increase robustness. Experiments verified that adversarial training enables the networks to withstand a variety of attacks. Moreover, it was confirmed that data augmentation helps improve robustness against image-based variations.

In traditional machine learning, physical knowledge is often leveraged to handcraft a feature extractor, whose results are then provided to a general function approximator. In deep learning, the temporal-data tensor can either be fully utilized as the input to 3D convolutional networks or restructured into a 4D tensor to allow the deployment of recurrent networks. Nonetheless, temporal redundancy is not trivially exploited, leading to a high overfitting risk when the training set is small. As such, attention-based architectures and transformer networks have been employed for noise-robust beamforming in order to scale to long-range dependencies within the temporal dimension. In addition to temporal resolution, dimensionality also matters from the perspective of modeling data distribution. With a fixed aperture size and number of training samples, increasing the array dimension leads to greater data redundancy, enhances modeling ability, and thus contributes to better generalization performance. Nevertheless, this consideration becomes less significant when training samples are abundant.



Equation 3: Delay-and-Sum (DAS) beamformer derivation

The article names delay-and-sum as a standard baseline.

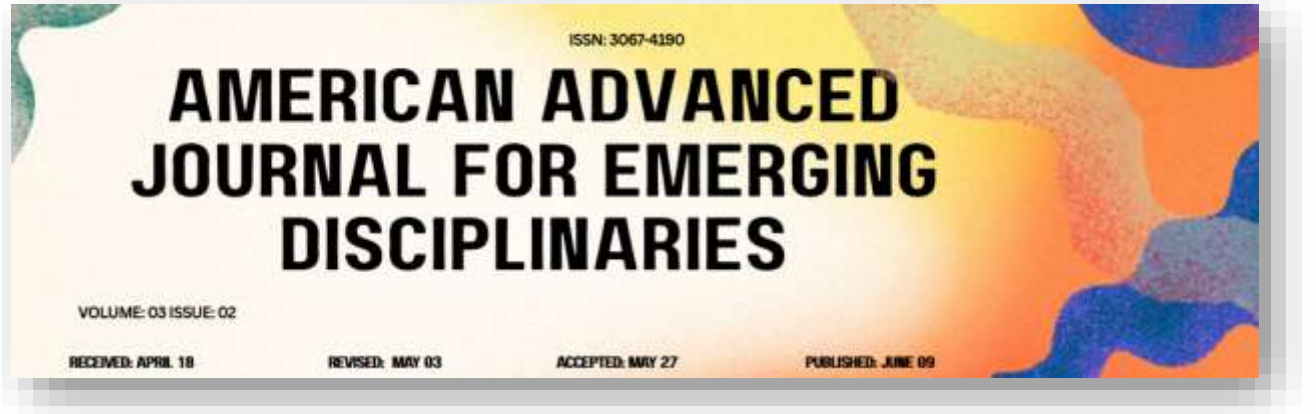
3.1 Constraint idea

We want to **phase-align** a wave from look direction θ_0 . Choose:

$$\mathbf{w}_{\text{DAS}}(\theta_0) = \frac{1}{M} \mathbf{a}(\theta_0)$$

3.2 Why this works (step-by-step)

If the true source is at θ_0 , $\mathbf{x} = \mathbf{a}(\theta_0)s + \mathbf{n}$, then:



$$y = \mathbf{w}^H \mathbf{x} = \left(\frac{1}{M} \mathbf{a}(\theta_0) \right)^H (\mathbf{a}(\theta_0)s + \mathbf{n}) \quad y = \frac{1}{M} \mathbf{a}(\theta_0)^H \mathbf{a}(\theta_0)s + \frac{1}{M} \mathbf{a}(\theta_0)^H \mathbf{n}$$

For typical steering-vector normalization, $\mathbf{a}^H \mathbf{a} = M$, so:

$$y \approx s + (\text{averaged noise term})$$

6.1. Mismatches, Noise, and Model Robustness

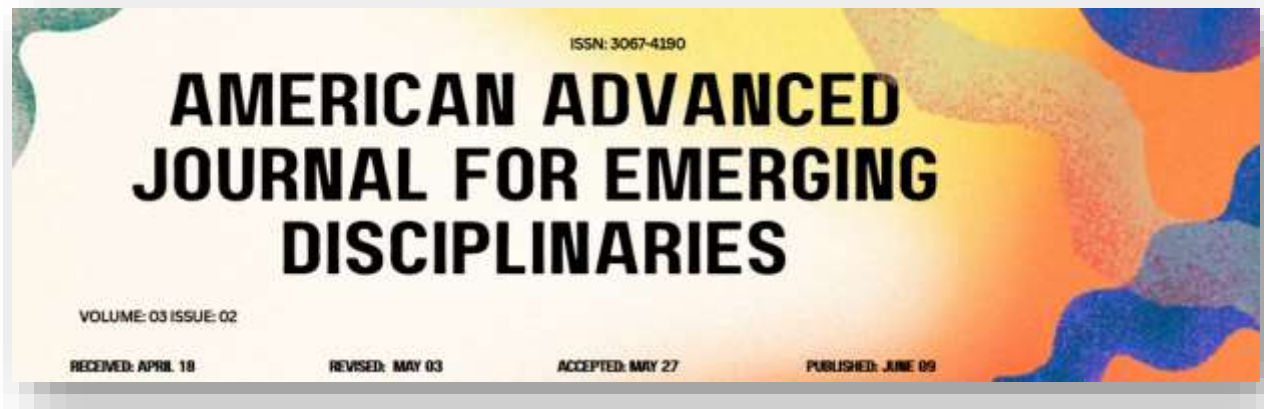
Diverse sources of training-data-model mismatch may severely undermine machine-learning (ML) performance despite successful training. Due to the sensor array's spatial structure, many applications involve limited amounts of training data, and often an even more stringent condition is met: all training samples come from a specific sensor configuration. Recognizing this model-data mismatch, the potential for model generalization beyond the data-support region has been analyzed in standard classification settings using simple Gaussian mixtures.

For example, a learned estimator can generalize beyond the configuration of a 2×2 linear array and to a 4×4 square aperture also not seen during training. For broadband sources, however, divergence of the model is expected when evaluating clash conditions for which the spatial covariance matrices are not positive semi-definite. The use of more complex class-condition distributions helps to alleviate this problem. Verification of a model's ability to generalize is a sufficient condition for robustness with respect to the adoption of a new configuration.

Dealing with noise is important since operating in an adversarial environment has become a security concern in several application domains. Providing resilience to such factors may even parallel an increase in the test-error rate. Mismatch in the channel response from the sensor array to the source of interest has also been addressed, although using a different data generation strategy.

7. Conclusion

Trained models address realistic but simplified data distributions; mismatches during testing can lead to poor generalization. Robustness to noise and channel distortion has been explored, but variability of stochastic channels remains largely unexamined. Evaluating adversarial attacks against trained models and devising defensive strategies is a critical but



seldom addressed direction. Unsupervised paradigms require fewer assumptions, enhance scaling to unlabeled data, and improve generalization performance by reducing reliance on an explicit data prior.

To date, most efforts targeting beamforming and direction-finding via ML focus on supervised schemes. However, supervised learning for imitation tasks can be entirely avoided if a set of good beamformers is first computed and subsequently exploited for unsupervised-training approaches. The magnitude of the output of each monopulse beamformer can be viewed as a good unsupervised representation of the input data. Using such representations to train the core of the radar or sonar system by finding the shrinkage beamformer parameters with an L1 loss is promising, as they allow for end-to-end training of a radar or sonar system where the monopulse beamformers are hidden. The next step to further investigate is to model the training process of any beamforming type in a self-supervised manner.

Contributions to Training Representations

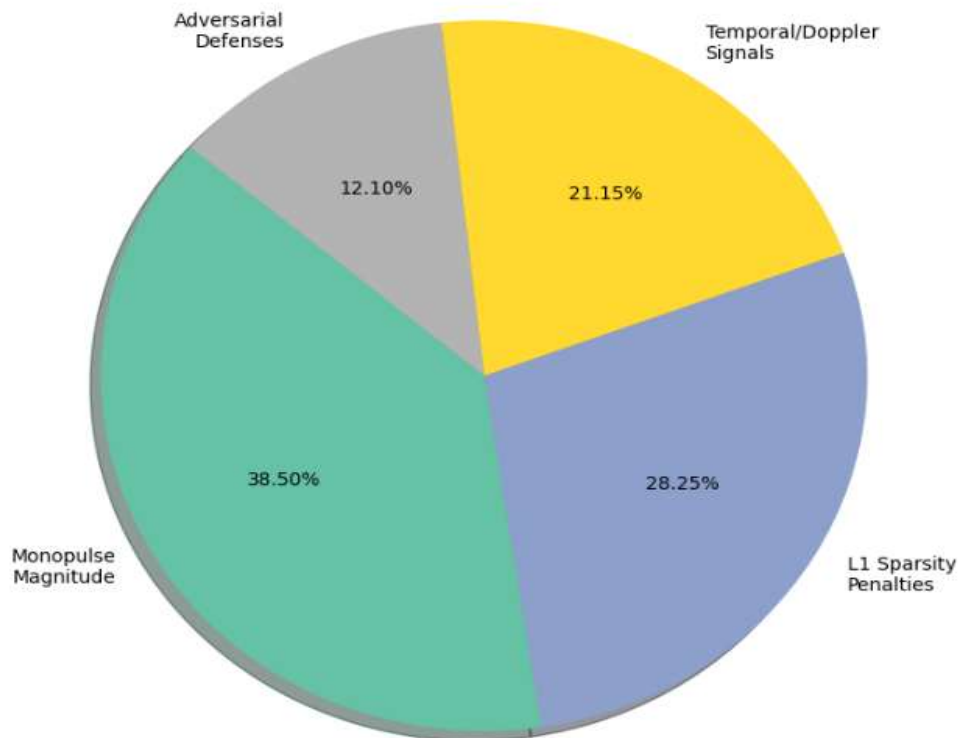
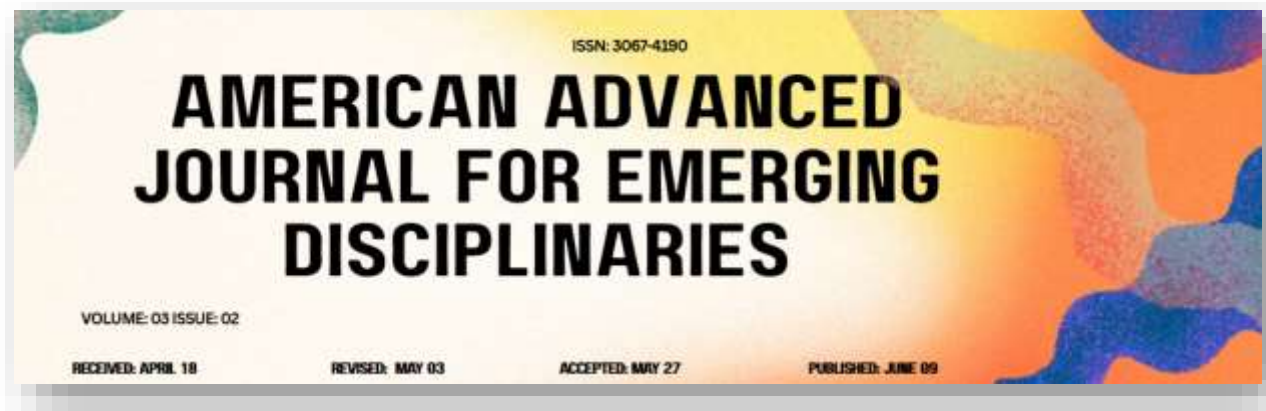


Fig 4: Contributions to Training Representations

7.1. Summary and Future Directions

Robustness and generalization of machine learning-based beamforming and direction finding methods remain active areas of investigation. The staged nature of the task introduces various possible sources of mismatch. A joint-end to end training approach provides the maximum level of performance, while self-supervised learning from the unlabeled temporally-contiguous signals of a sonar or radar system provides the most generalizable architecture. Both noise and channel variability are recognized to reduce performance, with the introduction of

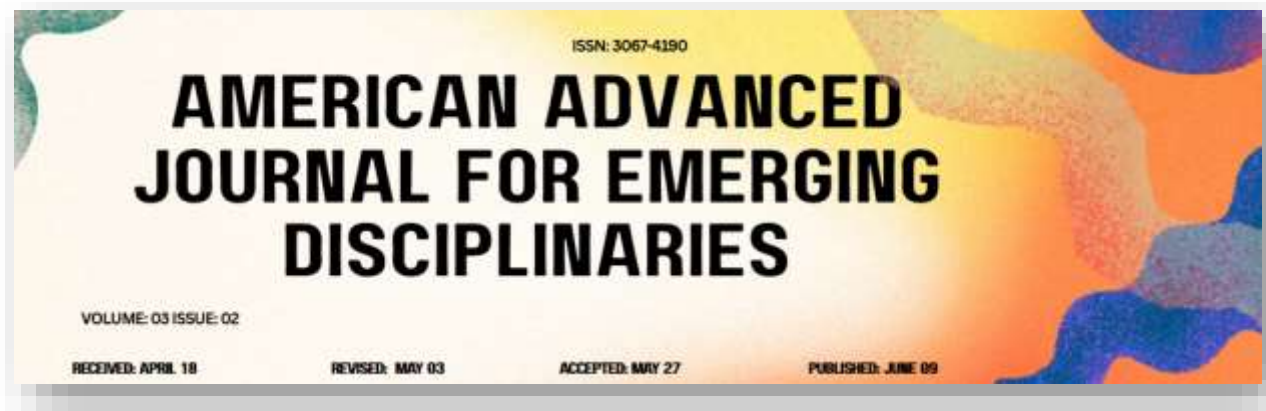


noise-robustness measures proving helpful in guiding the design of more capable networks. Exploration of adversarial machine learning in the spatial domain is at an early stage.

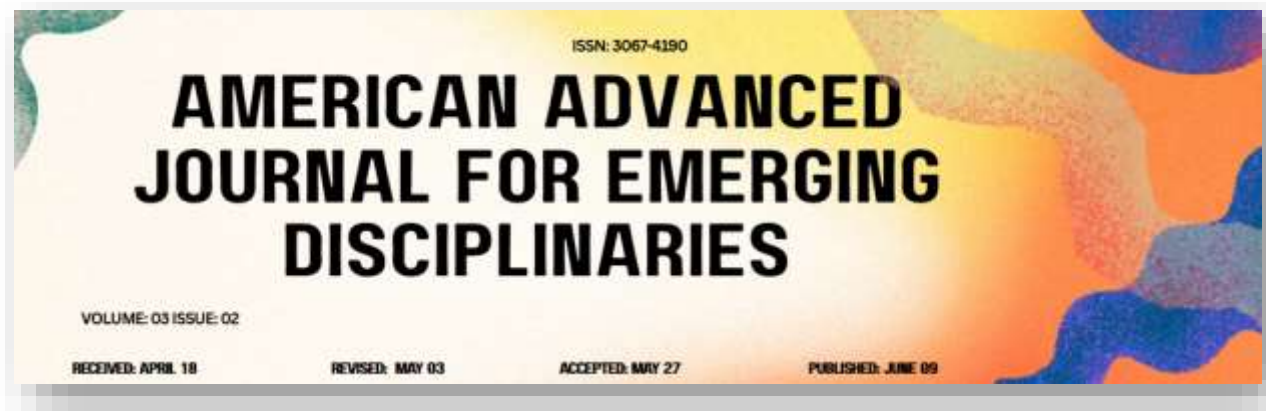
Several promising avenues of research remain open. Self-supervised methods still require the introduction of auxiliary labels in order to achieve state-of-the-art performance, as they are currently unsuitable for more than two classes. Their development for multi-class scenarios directly on the radar/sonar signals is therefore an important area for future exploration. Other natural extensions include the use of structured prediction and the incorporation of radar/sonar data in which the source distribution is unknown.

References

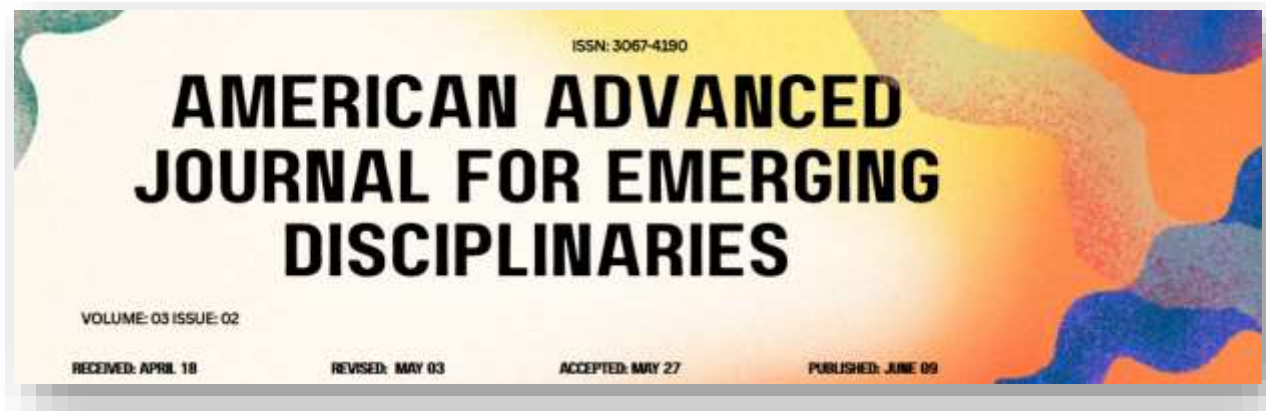
1. Barthelme, A., & Utschick, W. (2021). DoA estimation using neural network-based covariance matrix reconstruction. *IEEE Signal Processing Letters*, 28, 783–787.
2. Wandale, S., & Ichige, K. (2021). Design of sparse arrays via deep learning for enhanced DOA estimation. *EURASIP Journal on Advances in Signal Processing*, 2021, 17.
3. Lyu, X., Wang, J., & Zhang, Y. (2021). Direction of arrival estimation in passive radar based on deep neural network. *IET Signal Processing*, 15(8), 506–518.
4. Sriram, H. K., Challa, K., Gadi, A. L., & Singreddy, S. (2025). AI and Cloud-Driven Transformation in Finance, Insurance, and the Automotive Ecosystem: A Multi-Sectoral Framework for Credit Risk, Mobility Services, and Consumer Protection. Anil Lokesh and singreddy, Sneha, AI and Cloud-Driven Transformation in Finance, Insurance, and the Automotive Ecosystem: A Multi-Sectoral Framework for Credit Risk, Mobility Services, and Consumer Protection (March 15, 2025).
5. Liu, Y., Tong, F., Zhong, S., Hong, Q., & Li, L. (2021). Reverberation aware deep learning for environment tolerant microphone array DOA estimation. *Applied Acoustics*, 184, 108337.
6. Ge, S., Li, K., & Rum, S. N. B. M. (2021). Deep learning approach in DOA estimation: A systematic literature review. *Mobile Information Systems*, 2021, 6392875.
7. Seenu, A., Sheelam, G. K., Motamary, S., Meda, R., Koppolu, H. K. R., & Inala, R. (2025). AI-Driven Innovations in Infrastructure Management with 6G Technology. In 2025 2nd International Conference on Computing and Data Science (ICCDs) (pp. 1–6). IEEE. 2025 2nd International Conference on Computing and Data Science (ICCDs). <https://doi.org/10.1109/iccds64403.2025.11209649>



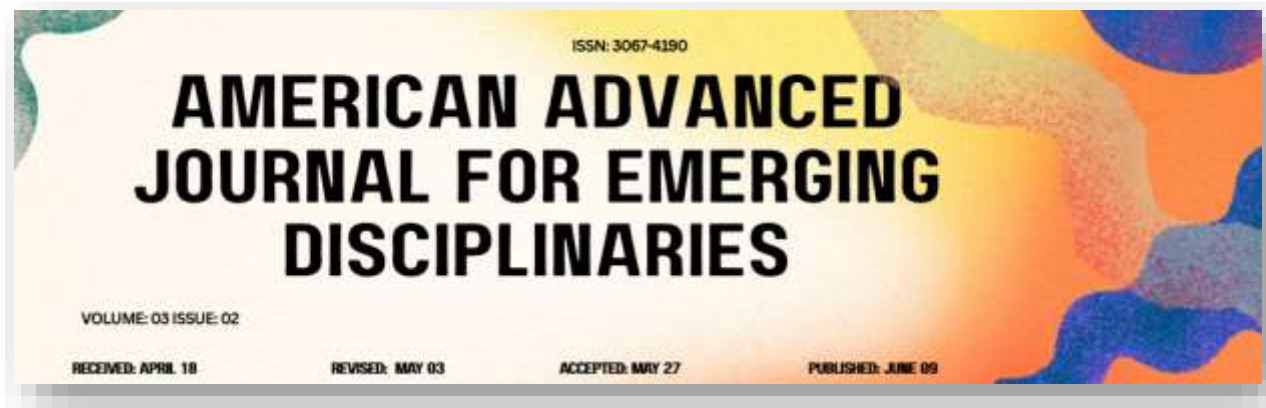
8. Huang, X., Bell, M. A. L., & Ding, K. (2021). Deep learning for ultrasound beamforming in flexible array transducer. *IEEE Transactions on Medical Imaging*, 40(11), 3178–3189.
9. Kim, H., Kim, J., & Lee, S. (2021). DoA estimation for FMCW radar by 3D-CNN. *Sensors*, 21(16), 5319.
10. Narasareddy Annapareddy, V., Pamisetty, A., Malempati, M., Kaulwar, P. K., & Bhardwaj Komaragiri, V. (2025, April). Enhancing Solar Power System Efficiency Through AI-Driven Predictive Maintenance and Cloud-Based Infrastructure Stability Solutions. In *International Conference on Smart Computing and Informatics* (pp. 328-337). Cham: Springer Nature Switzerland.
11. Jiang, W., Ren, Y., Liu, Y., & Leng, J. (2022). Artificial neural networks and deep learning techniques applied to radar target detection: A review. *Electronics*, 11(1), 156.
12. Chen, D., Shi, S., Gu, X., & Shim, B. (2022). Robust DoA estimation using denoising autoencoder and deep neural networks. *IEEE Access*, 10, 52551–52564.
13. Dutta, P., Adusupalli, B., Koppolu, H. K. R., Dodda, A., Yağanoğlu, M., Banerjee, J. S., & Chakraborty, A. (2025, March). Textual Social Data Disinformation Analysis Using a Hybrid Context-Enhanced Deep Learning Model. In *Doctoral Symposium on Human Centered Computing* (pp. 342-352). Singapore: Springer Nature Singapore.
14. Hoang, D. T., & Lee, K. (2022). Deep learning-aided coherent direction-of-arrival estimation with the FTMR algorithm. *IEEE Transactions on Signal Processing*, 70, 1118–1130.
15. Chung, H., Park, H., & Kim, S. (2022). Leveraging deep learning for practical DoA estimation: Experiments with real data collected via USRP. *Sensors*, 22(19), 7578.
16. Challa, S. R., Burugulla, J. K. R., Pamisetty, A., Challa, K., & Paleti, S. (2025, April). AI and ML-Powered Cybersecurity Strategies for Cloud Computing: Ensuring Infrastructure Stability in Financial and Retail Sectors. In *International Conference on Smart Computing and Informatics* (pp. 315-327). Cham: Springer Nature Switzerland.
17. Muthuchidambaranathan, P., Jeyakumar, P., Manikandan, E., & Muthukumaran, P. (2022). Beamforming design with fully connected analog beamformer using deep learning. *International Journal of Communication Systems*, 35(7), e5109.
18. Maden, K., & Erer, I. (2022). DOA estimation in MIMO radars via deep learning. In *Proceedings of the IEEE Radar Conference*.
19. Challa, S. R., Kaulwar, P. K., Koppolu, H. K. R., Adusupalli, B., & Suura, S. R. (2025, April). Big Data and AI in Wealth Management: Leveraging Cloud Computing for Secure and Scalable Financial Infrastructure. In *International Conference on Smart Computing and Informatics* (pp. 307-318). Cham: Springer Nature Switzerland.



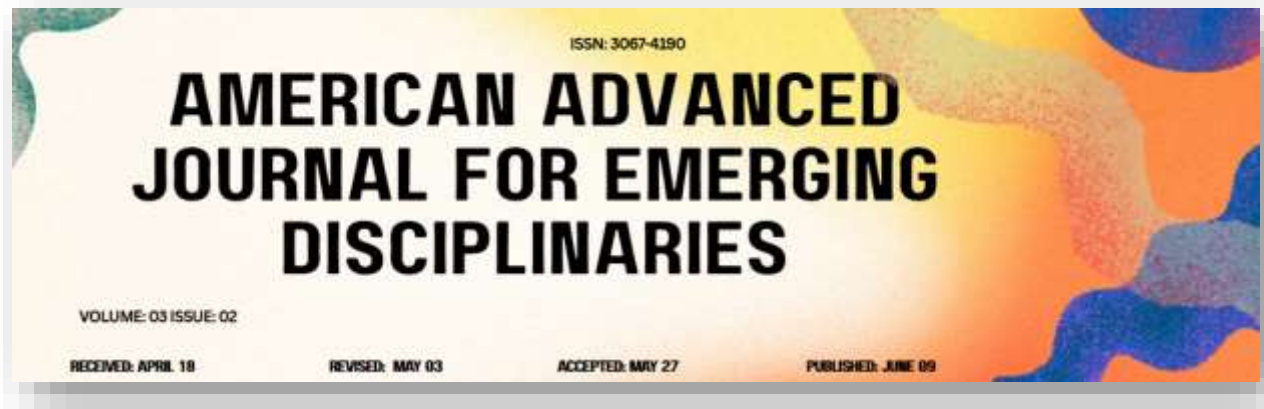
20. Kassir, H. A., Zaharis, Z. D., Lazaridis, P. I., Kantartzis, N. V., Yioultsis, T. V., & Xenos, T. D. (2022). A review of the state of the art and future challenges of deep learning-based beamforming. *IEEE Access*, 10, 80869–80882.
21. Chen, T., Shen, M., Guo, L., & Hu, X. (2023). A gridless DOA estimation algorithm based on unsupervised deep learning. *Digital Signal Processing*, 133, 103823.
22. Qin, Y. (2023). Deep networks for direction of arrival estimation with sparse prior in low SNR. *IEEE Access*, 11, 44637–44648.
23. Singireddy, J., Kaulwar, P. K., Somu, B., Meda, R., Dodda, A., & Yellanki, S. K. (2025, August). Reinforcement Learning-Based Asset Allocation in Algorithmic Trading for Banking Institutions. In *International Conference on Artificial Intelligence: Theory and Applications* (pp. 357-369). Cham: Springer Nature Switzerland.
24. Lan, X., Zhai, H., & Wang, Y. (2023). A novel DOA estimation of closely spaced sources using attention mechanism with conformal arrays. *IEEE Access*, 11, 44010–44018.
25. Li, Y., Shi, B., Shu, F., Song, Y., & Wang, J. (2023). Deep learning-based DOA estimation for hybrid massive MIMO receive array with overlapped subarrays. *EURASIP Journal on Advances in Signal Processing*, 2023, 110.
26. Davuluri, P. S. L. (2023). AI-Augmented Sanctions Screening: Enhancing Accuracy and Latency in Real Time Compliance Systems. *AI-Augmented Sanctions Screening: Enhancing Accuracy and Latency in Real Time Compliance Systems* (December 15, 2023).
27. Gao, S., Ma, H., Liu, H., Yang, J., & Yang, Y. (2023). A gridless DOA estimation method for sparse sensor array. *Remote Sensing*, 15(22), 5281.
28. Vadisetty, R., Nuka, S. T., Kalisetty, S., Pandugula, C., Burugulla, J. K. R., & Annapareddy, V. N. (2025, February). Generative AI for Advanced Recycling Processes in Polyethylene and Polypropylene Manufacturing. In *International Ethical Hacking Conference* (pp. 269-284). Singapore: Springer Nature Singapore.
29. Kowalk, U., Doclo, S., & Bitzer, J. (2023). Geometry-aware DoA estimation using a deep neural network with mixed-data input features. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*.
30. Subramanian, A. S., Weng, C., Watanabe, S., Yu, M., & Yu, D. (2022). Deep learning based multi-source localization with source splitting and its effectiveness in multi-talker speech recognition. *Computer Speech & Language*, 75, 101360.
31. Singireddy, S., Pandiri, L., Motamary, S., Kummari, D. N., Somu, B., & Lakkarasu, P. (2025, August). Blockchain-Powered Claims Validation for Enhancing Trust in Health and Auto Insurance Ecosystems. In *International Conference on Artificial Intelligence: Theory and Applications* (pp. 26-38). Cham: Springer Nature Switzerland.



32. Elbir, A. M., et al. (2023). Deep learning-based direction finding in the presence of direction-dependent mutual coupling. *ICT Express*, 9(4), 670–676.
33. Pandugula, C., Ganti, V. K. A. T., Kalisetty, S., Ande, R., & Mehta, D. (2025). Modeling the Som Kamla Amba Sub-Watershed Using Artificial Neural Networks (ANN) and the SWAT Tool. Available at SSRN 5341614.
34. Fadakar, A., Jafari, A., Tavana, P., Jahani, R., & Akhavan, S. (2024). Deep learning based 2D-DOA estimation using L-shaped arrays. *Journal of the Franklin Institute*, 361(6), 106743.
35. Zheng, S., Yang, Z., Shen, W., Zhang, L., et al. (2024). Deep learning-based DOA estimation. *IEEE Transactions on Cognitive Communications and Networking*, 10(3), 819–835.
36. Wang, M., Liu, X., Na, X., Zhang, Y., Liu, Y., & Qiu, T. (2024). Neural network-based DOA estimation for distributed sources in massive MIMO systems. *AEU—International Journal of Electronics and Communications*, 176, 155132.
37. KOTHAPALLI, S. L., SEENU, A., DILEEP, V., & YASMEEN, Z. (2025). THE FUTURE OF CUSTOMER ENGAGEMENT IN RETAIL BANKING: EXPLORING THE POTENTIAL OF AUGMENTED REALITY AND IMMERSIVE TECHNOLOGIES. *INTERNATIONAL JOURNAL*, 73(1), 72-79.
38. Chen, P., Chen, Z., Liu, L., Chen, Y., & Wang, X. (2024). SDOA-Net: An efficient deep-learning-based DOA estimation network for imperfect array. *IEEE Transactions on Instrumentation and Measurement*, 73, 8503512.
39. Zhang, B., He, J., Liu, P., Wang, L., et al. (2024). An automated multi-layer perceptron discriminative neural network based on Bayesian optimization achieves high-precision one-source single-snapshot direction-of-arrival estimation. *Scientific Reports*, 14, 10300.
40. Mylonakis, C. M., & Zaharis, Z. D. (2024). A novel three-dimensional direction-of-arrival estimation approach using a deep convolutional neural network. *IEEE Open Journal of Vehicular Technology*, 5, 643–657.
41. Ande, R., Mehta, D., Pandugula, C., Krishna AzithTejaGanti, V., & Kalisetty, S. (2025). Modeling the Som Kamla Amba sub-watershed using Artificial Neural Networks (ANN) and the SWAT tool. Available at SSRN 5341755.
42. Kassir, H. A., Kantartzis, N. V., Lazaridis, P. I., Sarigiannidis, P., Goudos, S. K., Christodoulou, C., & Zaharis, Z. D. (2024). Improving DOA estimation via an optimal deep residual neural network classifier on uniform linear arrays. *IEEE Open Journal of Antennas and Propagation*, 5(2), 460–473.



43. Liu, A., Guo, J., Arnatovich, Y., & Liu, Z. (2024). Lightweight deep neural network with data redundancy removal and regression for DOA estimation in sensor array. *Remote Sensing*, 16(8), 1423.
44. Bianco, S., Celona, L., Crotti, P., Napoletano, P., Petraglia, G., & Vinetti, P. (2024). Enhancing direction-of-arrival estimation with multi-task learning. *Sensors*, 24(22), 7390.
45. Lu, Y., Guan, H., Yang, K., Peng, T., Wen, C., & Li, X. (2024). Improving the accuracy of direction of arrival estimation with multiple signal inputs using deep learning. *Sensors*, 24(10), 2971.
46. Maguluri, K. K. (2025). Ethical challenges in artificial intelligence. *How Artificial Intelligence is Transforming Healthcare IT: Applications in Diagnostics, Treatment Planning, and Patient Monitoring*, 132.
47. Liu, L., et al. (2024). Direction of arrival estimation of coherent sources via a signal space deep convolution network. *Symmetry*, 16(4), 433.
48. Xu, X., & Huang, Q. (2024). MD-DOA: A model-based deep learning DOA estimation architecture. *IEEE Sensors Journal*.
49. Raiguru, P., Rout, S. K., Sahani, M., & Mishra, R. K. (2024). Machine learning-aided sparse direction of arrival estimation. *IEEE Sensors Journal*, 24(22), 38125–38134.
50. Su, Y., Wang, X., & Li, L. (2024). A novel off-grid deep learning framework for DOA estimation. In *Proceedings of the 2024 IEEE International Conference on Signal, Information and Data Processing*.
51. Monga, S., et al. (2024). Advanced beamforming and reflection control in intelligent reflecting surfaces with integrated channel estimation. *IET Microwaves, Antennas & Propagation*, 18(12), 917–931.
52. Di Barba, P., & Januszkiewicz, Ł. (2023). Linear antenna array modeling with deep neural networks. *Journal of Applied Electromagnetism*, 73(4).
53. Cousik, T. S., Shah, V. K., Reed, J. H., Tran, H. X., & Jana, R. (2023). Deep learning based fast and accurate beamforming for millimeter-wave systems. *IEEE Wireless Communications Letters*.
54. Hu, R., Jiang, L., & Li, P. (2021). Hybrid beamforming with deep learning for large-scale antenna arrays. *IEEE Access*, 9.
55. ADUSUPALLI, B. (2025). *Redefining Financial Risk Strategies: The Integration of Smart Automation, Secure Access Systems, and Predictive Intelligence in Insurance, Lending, and Asset Management*. JAIBDD.
56. Ma, K., Wang, Z., Tian, W., Chen, S., & Hanzo, L. (2021). Deep learning for beam-management: State-of-the-art, opportunities and challenges. *IEEE Wireless Communications*.



57. Jiang, W., et al. (2021). Deep-learning for radar: A survey. *IEEE Access*, 9, 141800–141818.
58. Bhalke, D., Paikrao, P. D., & Anguera, J. (2024). Deep learning-based beamforming approach incorporating linear antenna arrays. *Journal of Telecommunications and Information Technology*.
59. Ahmadinejad, A., & Talebi, S. (2025). Beamforming design via machine learning in intelligent reflecting surface-aided wireless communication. *Physical Communication*, 68, 102586.
60. Zhao, Y., Fan, X., Liu, J., Li, Y., Yao, L., & Wang, J. (2025). Efficient 2D-DOA estimation based on triple attention mechanism for L-shaped array. *Sensors*, 25(8), 2359.