Integrated Radio Sensing Capabilities for 6G Networks: AI/ML Perspective

(article starts on next page)

Integrated Radio Sensing Capabilities for 6G Networks: AI/ML Perspective

Victor Shatov, *Graduate Student Member, IEEE*, Steffen Schieler, *Graduate Student Member, IEEE*,
Charlotte Muth, *Graduate Student Member, IEEE*, José Miguel Mateos-Ramos, *Graduate Student Member, IEEE*,
Ivo Bizon, *Graduate Student Member, IEEE*, Florian Euchner, *Graduate Student Member, IEEE*,
Sebastian Semper, Stephan ten Brink, *Fellow, IEEE*, Gerhard Fettweis, *Fellow, IEEE*,
Christian Häger, *Member, IEEE*, Henk Wymeersch, *Fellow, IEEE*, Laurent Schmalen, *Fellow, IEEE*,
Reiner Thomä, *Life Fellow, IEEE*, and Norman Franchi, *Member, IEEE*

Abstract—The sixth-generation wireless communications (6G) is often labeled as "connected intelligence". Radio sensing, aligned with machine learning (ML) and artificial intelligence (AI), promises, among other benefits, breakthroughs in the system's ability to perceive the environment and effectively utilize this awareness. This article offers a panoramic view of radio sensing by unifying physical object sensing and spectrum sensing. To this end, while staying in the framework of integrated sensing and communication (ISAC), we expand the term "sensing" from radar, via spectrum sensing, to miscellaneous applications of radio sensing like non-cooperative transmitter localization. We formulate the problems, explain the state-of-the-art approaches, and detail AI-based techniques to tackle various objectives in the context of wireless sensing. Finally, we discuss the potential integration of various radio sensing functions into a common AI-enhanced framework, emphasizing the possible benefits and the challenges to overcome. In addition to the tutorial-style core of this work based on direct authors' involvement in 6G research problems, we review the related literature, and provide both a good start for those entering this field of research, and a topical overview for a general reader with a background in wireless communications.

Index Terms—6G, artificial intelligence (AI), channel charting, integrated sensing and communications (ISAC), localization, machine learning (ML), radar, signal classification, spectrum sensing, wireless sensing.

I. Introduction

It is generally accepted that wireless networks, once designed solely for transmitting voice calls, will continue to evolve, becoming significantly more capable and intelligent. Similarly to how the five human senses perceive raw data from our surroundings for subsequent processing and interpretation by the brain, radio sensors can feed the wireless network with diverse environmental data, enabling the network's capability to adapt and optimize autonomously. Nowadays, the research community is exploring data-driven solutions as promising alternatives to traditional algorithms based on domain knowledge, statistics and mathematical models.

A. Motivation and Background

Since around a decade ago, algorithms based on artificial intelligence (AI) and machine learning (ML) have prevailed in speech recognition, visual object recognition, and object detection [1], [2], [3]. Meanwhile, a relatively small number of enthusiasts tried to leverage ML-based techniques for wireless communications. Though ML was long envisioned as a natural way to adapt wireless networks based on spectrum sensing in the context of cognitive radio (CR) [4], [5], its potential for wireless communication systems was clearly and systematically articulated only in the second half of the 2010s [6], [7], [8]. Moreover, around that time, wireless networks were predicted to go well beyond mobile Internet and be ML-powered in early roadmaps for sixth-generation wireless communications (6G) [9]. As of 2026, ML-based algorithms have become mainstream in wireless research on the physical layer, while practical implementations are mostly still in their infancy.

Although the ML origins can be traced back to 1950s, the recent exponential advancement in data-driven methods is mainly due to the following three factors: (*i*) improvement in algorithms and model architectures; (*ii*) breakthrough in computational capabilities, notably with the development of graphics processing units (GPUs) and specialized hardware like Google's tensor processing units (TPUs); (*iii*) availability

Victor Shatov and Norman Franchi are with the Institute for Smart Electronics and Systems (LITES), Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, 91058, Germany (e-mail: victor.shatov@fau.de; norman.franchi@fau.de). Their work contributes to research within the 6G-Valley innovation cluster.

Steffen Schieler, Sebastian Semper, and Reiner Thomä are with the Electronic Measurements and Signal Processing Group, Technische Universitat Ilmenau, Germany (e-mail: steffen.schieler@tu-ilmenau.de; reiner.thomae@tu-ilmenau.de).

Charlotte Muth and Laurent Schmalen are with the Communications Engineering Lab (CEL), Karlsruhe Institute of Technology (KIT) Hertzstr. 16, 76187 Karlsruhe, Germany (e-mail: charlotte.muth@kit.edu; laurent.schmalen@kit.edu).

José Miguel Mateos-Ramos, Christian Häger, and Henk Wymeersch are with the Department of Electrical Engineering, Chalmers University of Technology, Sweden (e-mail: josemi@chalmers.se; christian.haeger@chalmers.se; henkw@chalmers.se).

Ivo Bizon and Gerhard Fettweis are with the Vodafone Chair Mobile Communications Systems, Technische Universität Dresden (TUD), Germany (e-mail: ivo.bizon@ifn.et.tu-dresden.de; fettweis@ifn.et.tu-dresden.de).

Florian Euchner and Stephan ten Brink are with the Institute of Telecommunications, University of Stuttgart, 70569 Stuttgart, Germany (e-mail: florian.euchner@inue.uni-stuttgart.de; tenbrink@inue.uni-stuttgart.de).
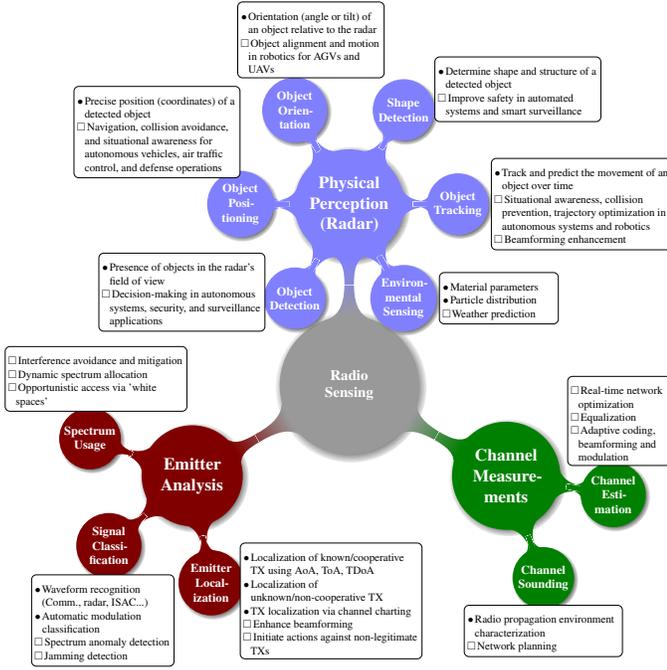
Fig. 1. A taxonomy of sensing capabilities discussed in this article, along with some of their potential applications (□). In ISAC, physical perception and channel measurements are typically obtained with cooperating TXs and RXs, while emitter analysis is often non-cooperative.

of huge amounts of data. In the wireless domain, (*i*) and (*ii*) can be expected and attributed to natural technological progress over time, while (*iii*) seems to be the bottleneck. Compared to audio or image signals, high quality wireless raw data is more difficult to obtain. For instance, such data frames might need to include details like interference patterns, channel conditions, and user behavior, which are not trivial to collect or label in plenty. The data shortage problem is understood and currently being tackled by the 3rd Generation Partnership Project (3GPP). In particular, enhanced data collection is part of the recently frozen 3GPP Release 18 [10].

Sensing ubiquitous wireless signals will facilitate the extraction of abundant data and help to achieve complete environmental awareness. Following [11], we roughly categorize radio sensing as follows: (*i*) radar sensing to extract the information about the physical world; (*ii*) channel measurements to understand the propagation characteristics and parameters, and (*iii*) electromagnetic (EM) analysis of the environment, or spectrum monitoring. Fig. 1 illustrates this classification with a few details on each type of radio sensing. In this work, we group the first two as spatial awareness enablers with their own transmitter (TX), with the underlying goal of extracting information regarding physical objects. In addition, objects are assumed to be passive, i.e., they do not have any capability to transmit signals themselves. In contrast, the aim of the last category is to obtain EM awareness by identifying who is transmitting and where. EM awareness is purely passive and mostly non-cooperative.

The consideration of physical object sensing and spectrum analysis together stems from (*i*) their shared physical foundations and (*ii*) the growing need for efficient, intelligent, and adaptive wireless systems. Traditionally treated as separate domains, these sensing paradigms actually rely on analyzing

radio wave interactions with the environment, whether to detect objects or to characterize spectral occupancy. Furthermore, technological convergence is already occurring at the hardware level, as modern wireless devices can support both spectrum analysis and physical environment sensing using the same radio frequency (RF) front-ends [12]. Unifying these approaches enables more holistic environmental awareness, reduces hardware redundancy, and aligns with the vision of future wireless networks, where communications and different types of radio sensing are integrated into a network infrastructure.

As a concrete use case example, consider smart factory environments, where automated guided vehicles (AGVs) rely on radar sensing to detect obstacles and ensure safe navigation among machinery and workers [13]. Simultaneously, spectrum sensing unit(s), which might be mounted on AGVs, exemplifying so-called ego-view, or be a part of the factory infrastructure potentially distributed in the area, monitor the wireless environment for interference from malfunctioning devices or unauthorized emitters operating in shared bands. By combining both radio sensing modalities, the system can not only identify physical hazards and signal disruptions, but also correlate them, e.g., linking a moving interferer to a specific object. This enables proactive responses such as frequency adaptation, power control, or triggering maintenance. Signal classification can subsequently distinguish between benign and harmful interference, enabling anomaly detection. Unlike isolated sensing approaches, the unified method improves situational awareness, communication reliability, and safety, which are among the key requirements for Industry 4.0 and future 6G-enabled industrial automation [14].

Processing large amounts of diverse sensing data is challenging, and mathematical models often cannot capture the dependencies that may occur due to their dynamic nature or the peculiarities of an environment. ML approaches can be better than algorithms based on mathematical models, particularly in the context of radio sensing, due to the following:

- *Model mismatch* and oversimplification: Model-based algorithms are based on idealized assumptions, such as Gaussian noise and linearity. In real RF environments, these assumptions are often not valid, leading to poor performance under complex conditions.
- *Scalability and high dimensionality:* Analytical models struggle to scale with high-dimensional inputs, such as raw IQ samples and multi-modal data. In contrast, ML models can learn directly from large, complex datasets without requiring tractable mathematical models.
- *Unknown or time-varying environments*: Traditional algorithms are typically static or rule-based and fail to adapt to dynamic changes (e.g., mobility or hardware variability). In contrast, ML can adapt or be retrained online to maintain performance in evolving scenarios.
- *Lack of good explicit models*: In many applications (e.g., interference classification, anomaly detection), the available analytical models are not reliable or require manual work which is impractical [15]. ML excels in such cases by learning patterns directly from data without requiring expert knowledge.

Therefore, it is worth noting that via radio sensing, the increased number of devices and overall complexity in wireless environments can be turned from a challenge into an opportunity. For example, localization algorithms might exploit signals broadcasted by TXs and captured by RXs to find the coordinates of objects or emitters. At the same time, ML-based algorithms can refine this process, increasing accuracy and robustness in complex environments. Furthermore, expansion of available frequency bands for 6G have been under consideration.[1] This added diversity can also be used to better understand the propagation environment by leveraging multi-band channel sounding, which provides a frequency-dependent view of how radio waves interact with their surroundings.

Despite potential advantages of ML-based algorithms for radio sensing, their design faces common challenges, which can be summarized as follows:

- Scarcity of labeled data, particularly captured over the air.
- Limited explainability of deep learning models.
- Need for generalization to unseen environments.
- Integration of domain knowledge, e.g., signal structure, propagation structure, statistical features, into learning pipelines.
- Diverse computational resources available at network devices, ranging from memory-constrained embedded systems to GPU-accelerated platforms.

These challenges are relevant to both spectrum and object sensing applications and further motivate the need for a unified perspective.

3GPP is examining the integration of ML from the 5G foundation to the emergence of 6G, which is promised to be the first AI-native generation [17], [18], [19]. In Release 18, finalized in 2024, 3GPP, for the first time, investigated the potential of AI for a 5G air interface, while the ongoing Release 19, scheduled for completion at the end of 2025, continues this work with the focus on beam management and positioning. In parallel, 6G is envisioned to be the first generation that enables ISAC technology by both 3GPP [20] and International Telecommunication Union (ITU) [21], with ML being a natural choice to tackle various associated problems [22]. It is worth mentioning that spectrum sensing and analysis have received much less attention from standardization bodies, compared to radar-like sensing. In this context, the IEEE 802.15.22.3 standard seeks to unify the fragmented landscape of proprietary spectrum sensing and measurement systems by introducing a broadly applicable framework called "Spectrum Characterization and Occupancy Sensing" [23]; however, at the moment the working group activities seems to be on hold [24]. Enhanced with ML, radio sensing will provide a foundation towards building a digital twin (DT) of the radio environment [25], where the complete information about the physical and electromagnetic environment can serve

as a foundation for automated communication performance optimization, spectrum management, detection and localization of threats, as well as unveil new use cases for customers. Finally, yet equally significant, the trend towards ML in wireless networks is accepted and driven by traditional telecommunication stakeholders from all over the globe like Nokia [26], Ericsson [27], and Huawei [28]. At this point, the topic has attracted significant attention and investments from other domain big tech players like Nvidia [29], prominent for accelerated computing hardware, currently also offering tools for designing, simulating, training, and deploying AI-based wireless networks [30]. The recent announcement of a strategic partnership between Nvidia and Nokia, focused on integrating AI into the next generation of networks, marks the beginning of a new era in the telecommunications industry [31].

B. Related Works and Key Contributions

In the context of 6G, integrating sensing and ML capabilities into wireless networks is one of the hottest research topics, regardless of whether sensing and ML are considered together or separately. Here, we first familiarize the reader with key and recent works that fall into the review, tutorial, or perspective article categories. Then, we highlight the main contributions and novelty of this paper.

1) Focus on (passive) spectrum sensing: Two early surveys of ML for spectrum sensing in the context of CR, a concept formalized in the late 1990s, can be found in [32] and [33]. In [32], the authors formulate the ML-empowered CR observation through spectrum sensing with subsequent system reconfiguration and learning the impact of these actions on the performance of the radio. While [32] focuses on spectrum sensing to implement a practical CR and the relative merits of various proposed techniques in different CR applications, a later study [33] addresses supervised and unsupervised ML algorithms for both spectrum sensing and signal classification. A recent work [38] provided a survey on ML-driven CR wireless networks, including Internet of Things (IoT), mobile (namely vehicular and railway), and unmanned aerial vehicle (UAV) networks. This comprehensive survey also goes beyond spectrum occupancy detection techniques, reviewing various aspects of non-cooperative TX signal classification, which is denoted here as 'intelligent spectrum sensing'. A review of ML-driven jamming detection and interference management presents a slightly different angle of view on spectrum sensing in [39], where the authors presented a detailed taxonomy of intentional and unintentional spectrum-related troubles and ML-based mitigation techniques in various wireless communication scenarios, including proper resource allocation through spectrum sensing. In [42] and [43], surveys on ML methods for standard spectrum sensing, sharing, and access are provided. The authors of both reviews treat sensing as part of the classical CR technology scenario, in which secondary users detect the unutilized bands through sensing, thus facilitating better spectrum utilization. The paper [45] overviews AI in the context of multiple access for 6G, with spectrum sensing issues discussed along with protocol design and protocol optimization. The paper also briefly mentions a few works on automatic modulation classification (AMC) without going into detail.

---

[1]The bands available for 5G, which are also expected to be used in 6G, are categorized into two distinct frequency ranges (FR), where FR1 is from 410 MHz to 7.125 GHz, and FR2 from 24.25 to 52.6 GHz. The new FR3 , which spans the range between 7.125 and 24.25 GHz is under consideration since 2023 [16].

TABLE I

COMPARISON OF RELEVANT SURVEYS AND TUTORIALS ON SENSING FOR WIRELESS NETWORKS. SYMBOLS ✓ AND ✔ DENOTE PARTIALLY AND DETAILED COVERED TOPICS, RESPECTIVELY. CHECKMARKS IN THE "WIRELESS" AND "AI/ML" COLUMNS MARK WHETHER THE CORRESPONDING PAPER HAS A DEDICATED (SUB)SECTION WHERE SOME THEORETICAL BACKGROUND ON THE TOPIC IS PROVIDED. LOZENGE SYMBOL INDICATES WHETHER THE ARTICLE DISCUSSES LEVERAGING AI FOR A CERTAIN TASK, WHERE ◊ AND ♦ DENOTE A GLANCE (E.G., MENTIONING ONLY USE CASES AND OPPORTUNITIES) AND STRONG FOCUS (E.G., OVERVIEW OF CONCRETE AI/ML ALGORITHMS), RESPECTIVELY.

Ref.	Year	Key contributions	Basics		Physical perception of objects					Emitter analysis			Vision Aspect B5G
			Wireless	AI/ML	Object detection	Object positioning	Object orientation	Shape detection	Object tracking	Spectrum sensing	Signal classification	Transmitter Localization	
[32]	2010	One of the first known surveys with a clear focus on AI for CR		✓						✓♦			
[33]	2013	Overviews supervised and unsupervised ML techniques for both spectrum sensing and signal classification		✓						✓♦	✓♦		
[34]	2021	Identifies key localization and sensing enablers for B5G			✓◊	✓◊	✓◊	✓◊	✓◊			✓◊	✓
[35]	2022	Historical overview, performance trade-offs, waveform and receiver design of ISAC systems	✓		✓	✓			✓		✓◊		✓
[36]	2023	Focus on a vision of ML roles for ISAC in 6G			✓◊	✓◊	✓◊						✓
[13]	2024	Architectures, use cases, signal processing and hardware aspects of ISAC for B5G			✓	✓							✓
[37]	2024	Overview of the principles in ML and DL and their potential applications in ISAC, ML-enhanced ISAC		✓	✓◊	✓◊		✓◊	✓◊			✓◊	✓
[38]	2024	Comprehensive survey on ML-driven cognitive radio for wireless networks		✓						✓♦	✓♦		✓
[39]	2024	Review of AI/ML-driven jamming detection and interference management	✓	✓						✓♦	✓♦		✓
[40]	2024	Signal processing, applications and vision of ISAC systems	✓		✓◊	✓◊	✓◊	✓◊	✓◊			✓◊	✓
[41]	2024	Discussion on synchronization and overview of the other problems of ISAC	✓		✓◊	✓◊			✓				✓
[42]	2023	A survey on ML for spectrum sensing and signal classification		✓						✓♦	✓♦		
[43]	2024	Comprehensive survey on ML for spectrum sensing, sharing, and access		✓						✓♦			✓
[22]	2024	Bird's eye overview of AI-enhanced ISAC, including key challenges and perspectives			✓◊	✓◊	✓◊						✓
[44]	2024	Discussion on multimodal sensing and computation convergence in 6G		✓	✓♦	✓♦	✓♦		✓♦				✓
[45]	2024	AI for spectrum sensing in the context of multiple access for 6G		✓						✓♦	✓◊		✓
[46]	2025	Highlights principles, architecture, and potential applications of distributed ISAC (DISAC)			✓◊	✓◊	✓◊	✓◊	✓◊			✓	✓
[47]	2025	Up-to-date standardization activities, prototypes and testbeds for ISAC			✓◊	✓◊	✓◊	✓◊	✓◊				✓
[48]	2025	A broad survey of ML for ISAC in 6G also reviewing radar-communications coexistence	✓	✓	✓♦	✓♦	✓♦	✓♦	✓♦		✓◊	✓◊	✓
Our work	2026	A tutorial on various aspects of AI-Enhanced radio sensing for 6G	✓	✓	✓♦	✓♦	✓♦	✓♦	✓♦	✓♦	✓♦	✓♦	✓

2) Focus on (active) radar sensing: Narrowing down the term sensing purely to radar, [35] reviewed the historical development of ISAC, its performance gains and trade-offs, signal processing, potential waveforms, and the RX design. Interestingly, as one of the open challenges, the authors mentioned the importance of ML-based signal recognition in the context of radar-communication convergent systems. Similarly, the perspective paper [13] considers only radar sensing integration into 6G networks. The authors introduced the architectures of ISAC and reviewed use cases, waveforms, signal processing, and hardware aspects of ISAC, only mentioning ML as a potential performance enhancer. Another recent milestone work [40] provides a detailed overview and vision of ISAC in cellular networks, focusing on signal processing aspects for localization of sources and passive targets. Furthermore, this article discusses sensing-aided communications, with a focus on beam training for overhead reduction and blockage prediction in practical scenarios. In [41], additionally to the state-of-the-art ISAC overview as of 2024, the authors provided design guidelines and ten open challenges for ISAC systems. An early pre-standardization vision of 6G networks as convergent communication, localization, and sensing systems is formulated in [34]. That paper identifies key localization and sensing enablers for beyond 5G (B5G). This work focuses on futuristic integrated radar sensing solutions with fine-range, Doppler, and angular resolutions. It also notes the vital role of the flexible use of the RF spectrum and AI in 6G. A more recent study [49] took a step further and highlighted the importance of distributed observations of the environment, discussing principles, architecture, and potential applications of the distributed ISAC (DISAC). The survey [47] focuses on sensing of targets with an up-to-date overview

of standardization activities and testbeds. It also correctly mentions the need for spectrum sensing as a solution to accurately identify and effectively mitigate interference in [47, Sec. VIII.A]. ML-based approaches are only mentioned among others. On the other hand, another recent survey [48] is specifically dedicated to ML-enhanced ISAC. Since this survey contains a subchapter that reviews works on communication-radar coexistence, we consider it to be the closest to our paper in its scope. However, [48] is predominantly radar-dominated.

While some papers mention the value of the convergence of various sensing capabilities in 6G, their focus is often limited to only one modality. Hence, there is a need for a broader perspective on the different modalities of radio sensing. To the best of our knowledge, no reviews consider sensing of physical objects and EM sensing together. Meanwhile, to realize the expectations of 6G, e.g., build a DT of the wireless network, knowledge of both physical and EM worlds is equally essential. The main objective of this article is to present a holistic overview of various radio sensing aspects and open research and implementation problems, providing pointers towards AI-native 6G wireless systems. In this context, our paper aims not only to serve as a tutorial and survey, but also to act as a bridge between two radio sensing communities that at the moment seem to be quite separated, potentially leading to novel research. The key contributions can be summarized as follows:

- Since we view early career researchers as the primary audience of our article, we provide the reader with a comprehensive background, covering fundamental concepts of wireless channels and ML. In addition, we discuss the role of ray tracers and over-the-air measurements to collect wireless data, metrics to evaluate sensing performance, and different ways to represent and process data within modern ML frameworks.
- Based on our practical experience, we offer guidance on ML-enhanced radar sensing for the perception of physical objects with its possible integration into wireless networks. Widely regarded as one of the cornerstones in 6G, this new capability is presented from fundamental trade-offs to customer applications such as gesture recognition. The main focus is, however, on ML-enhanced target detection and parameter estimation.
- Similarly, we formulate the problems in the context of what we roughly categorize as passive/opportunistic emitter analysis, including CR-like standard spectrum sensing to detect underutilized parts of the spectrum and waveform classification, while providing ML-based existing solutions. Additionally, we provide unique insights on ML-empowered TX localization and channel charting, which is a self-supervised learning technique that can be used to facilitate non-cooperative TX localization.
- Finally, our paper summarizes previous efforts and considers the integration of *various* radio sensing functions enhanced by AI algorithms into future wireless networks, highlighting potential advantages and challenges to overcome. In this context, we discuss the hardware and architectural challenges of implementing such an

TABLE II
GLOSSARY: NOTATION

Variable	Explanation
t	Time (slow-time)
τ, τ_ℓ	Delay (fast-time)
f_c	Carrier frequency
L, with $\ell \in \{0, 1, \ldots, L-1\}$	Number of propagation paths
α_ℓ	Complex path weight
f_D	Doppler shift
$h(t, \tau), h_i(t, \tau)$	Channel impulse response
B_{f_D}	Doppler spread
T_s	Sampling time
f_s	Sampling frequency
T_c	Coherence time
s_{Tx}	Transmit signal
s_{Rx}	Received signal
$n(t)$	Noise process
P_D	Detection probability
$T, \hat{T}$	True and estimated number of targets
$\mathcal{P} = \{\boldsymbol{p}_1, ..., \boldsymbol{p}_{|\mathcal{P}|}\}$	True target positions
$\hat{\mathcal{P}} = \{\hat{\boldsymbol{p}}_1, ..., \hat{\boldsymbol{p}}_{|\hat{\mathcal{P}}|}\}$	Estimated target positions
$\mathcal{D}$	Dataset
$\boldsymbol{D} \in \mathcal{D}$	Sample of dataset
$\boldsymbol{\Theta}$	Trainable parameters of ML algorithm
$T(\boldsymbol{D})$	Task (e.g. class of dataset sample)
$f(\boldsymbol{D}, \boldsymbol{\Theta})$	Trainable function
$P(f(\boldsymbol{D}, \boldsymbol{\Theta}), T(\boldsymbol{D}))$	Performance measure

AI-powered system capable of simultaneously sensing physical and EM environments and communicating data.

Table I structures the scope of this work and compares it with other existing surveys, tutorials, and vision articles. This table focuses on recent articles, while a few selected older publications are introduced for completeness.

C. Organization of the Paper

Fig. 2 illustrates the article's organization. In Section II, we introduce the background in AI and wireless basics that make it easier for a general reader to follow the content of this work. Section III provides guidance to the AI-empowered physical perception of passive objects, focusing on radar capabilities integrated into the wireless network. Section IV aims to help navigate in the broad area of passive radio sensing, including spectrum sensing, signal classification, and TX localization. Section V highlights the importance of different types of radio sensing for the envisioned 6G wireless applications, discussing potential advantages and challenges. Finally, Section VI concludes the article with final remarks.

II. PRELIMINARIES

In this section, we first review the fundamental concepts and analytical expressions that are vital for understanding different types of wireless systems. Then, we summarize some ML basics and related terminology used throughout the article. Readers familiar with these preliminaries or those more interested in the overview can skip this section.

A. Band-Limited Wireless Channel Model

Knowledge of the channel is crucial in wireless networks because it enables efficient transmission and reception of signals, as well as effective use of resources. For communication systems, accurate channel estimation is essential to mitigate the effects of fading and multipath propagation. Furthermore,
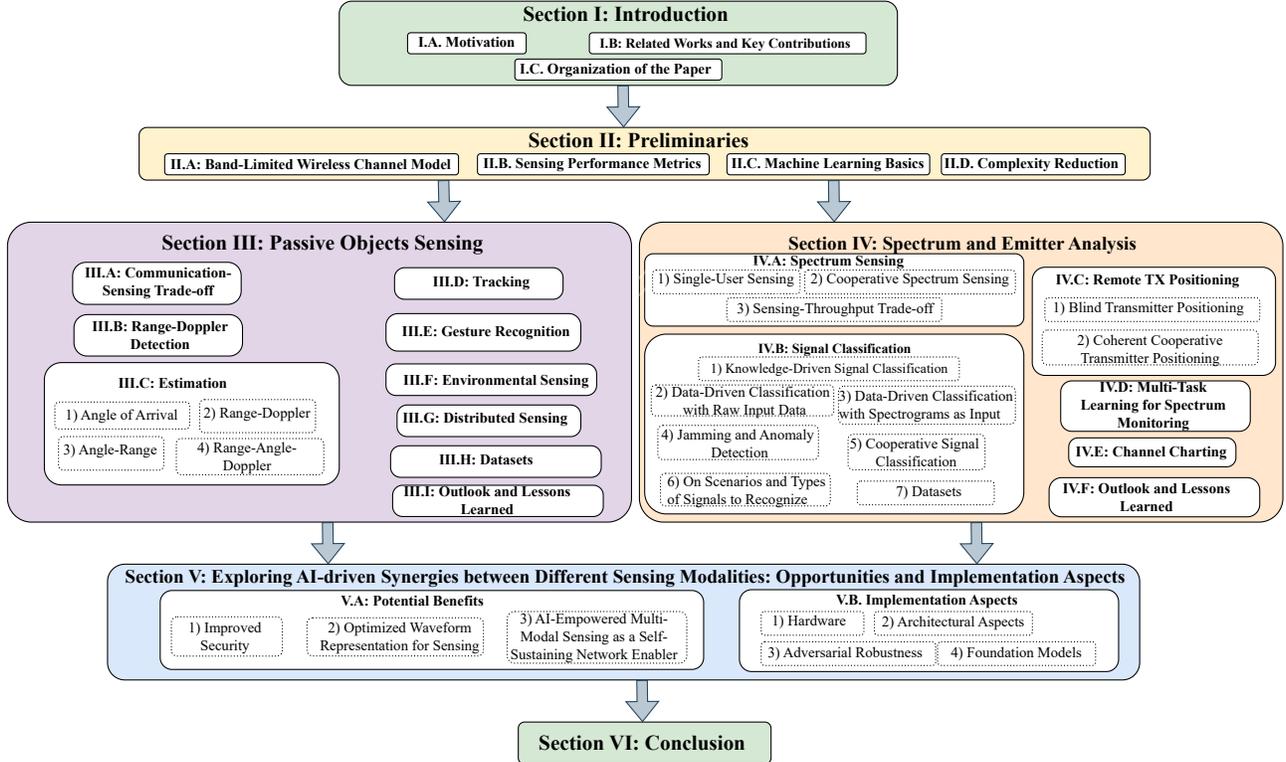
Fig. 2. The organization of the article.

since the channel estimate provides valuable information on how radio waves interact with the environment, its knowledge is a critical component in radio sensing. For example, it helps to detect, track, and characterize objects, as well as map the environment. Various signal processing algorithms are based on this information. Since real-world systems can only transmit signals within a specific frequency range defined by hardware and regulations, wireless channels are typically limited in bandwidth.

The channel impulse response (CIR) $h(t, \tau)$ describes how a single impulse at time τ is affected by the channel. When the transmitted signal is $s_{\text{Tx}}(t)$, the received signal $s_{\text{Rx}}(t)$ is given by

$$s_{\text{Rx}}(t) = \int_{-\infty}^{+\infty} s_{\text{Tx}}(\tau) h(t, \tau) d\tau + n(t), \qquad (1)$$

where $n(t) \sim \mathcal{CN}(0, N_0)$ is a Gaussian process corresponding to the additive white Gaussian noise (AWGN) with the noise power spectral density N_0.[2]

The received signal in (1) is the weighted sum of all time instances of the transmit signal. The weights are caused by multipath propagation in the environment, including signal reflection, diffraction and scattering.

The wireless propagation channel can be modeled as a superposition of incoming signal components that travel through multiple propagation paths. Therefore, they arrive at the RX with different delays, amplitudes, and phases. The

[2]Noise is often assumed to be AWGN after sampling and filtering.

combination of the signal components at the receiving antenna results in a signal that varies extensively both in amplitude and phase. Moreover, each signal component experiences a frequency shift that arises due to the relative motion between transmitter and receiver or moving reflectors, i.e., the Doppler shift. The complex baseband CIR at time t, with relative channel delay τ, after mixing the signal down from the carrier frequency f_c is expressed by

$$h(t, \tau) = \sum_{\ell=0}^{L(t)-1} \alpha_\ell(t) \delta\left(\tau - \tau_\ell(t)\right) e^{-j2\pi f_{\text{D},\ell} t}, \qquad (2)$$

where $L(t)$ represents the number of propagation paths, $\alpha_\ell(t)$, $\tau_\ell(t)$ represent the complex path weight and delay associated with the ℓ-th path, respectively, $f_{\text{D},\ell}$ denotes the Doppler shift, and $\delta(\cdot)$ the Dirac delta function [50], [51], [52]. To get a better insight into (2), let us imagine a TX in a room with an RX as a visual aid. The signal is transmitted directly through the line of sight (LoS) component, but also reflected by the walls. Through the different geometrical paths, the reflections arrive with a delay τ_ℓ, experience additional loss $|\alpha_\ell|$ and can be shifted in frequency by a Doppler shift $f_{\text{D},\ell}$ if any relative movement is lengthening or shortening the propagation path.

Equation (2) models the channels as a function of two time variables by assuming the channel is underspread such that $B_{f_D} \cdot \tau_{\text{rms}} \ll 1$, where B_{f_D} is the Doppler spread, which quantifies how much the received signal is spread in frequency due to motion, such that the Doppler shift changes only with t. The two time variables are commonly referred to as fast-time τ and slow-time t in radar sensing.

In digital receivers, the channel is sampled at a given sampling frequency $f_{\mathrm{s}} = 1/T_{\mathrm{s}}$ Hz, where T_{s} is the sampling time interval in seconds. When the difference among delays from propagation paths is smaller than T_{s}, these are vectorially combined into a single tap. The time dependency t can be dropped if the symbol time is smaller than the coherence time T_{c}. During the coherence time, channel responses are highly correlated, and the desired correlation level in a given time interval defines the relation between the coherence time T_{c} and the maximum Doppler shift $f_{\mathrm{D\,max}}$.

Lastly, we introduce the impulse response in antenna arrays that enable fundamental ISAC tasks such as beamforming and aiding radar localization through angle of arrival (AoA) estimation. If multiple RXs antenna elements are present, and they are spaced closely together, we can assume that similar signal components are received by all RXs. There is a small residual time delay between the different receiver nodes for each different path which can be modeled by rephrasing (2) for an antenna element i as

$$h_i(\tau) = \sum_{\ell=0}^{L-1} \alpha_\ell \delta\left(\tau - \tau_\ell - \tau_{i,\ell}\right) e^{j\Delta\varphi_{i,\ell}} \qquad (3)$$

with an additional geometric time delay $\tau_{i,\ell}$ between antenna i and a reference point. This extra delay caused a phase shift $\Delta\varphi_{i,\ell}$ depending on the delay difference with respect to the antenna elements.

B. Sensing Performance Metrics

Once this information is extracted from the CIR, the sensing performance can be evaluated using different metrics, depending on the goal of the sensing application. In the following, we describe the most common sensing performance evaluation metrics.

1) Detection probability: The detection probability is associated with the event that a target is detected when there is a target in the environment. In the single-target case, if $\hat{T} \in \{0,1\}$ is the estimated target appearance, where $\hat{T} = 0$ means that we do not detect a target and vice versa; and $T \in \{0,1\}$ is the true target appearance, then $P_{\mathrm{D}} = p(\hat{t} = 1 | t = 1)$[3].

2) False alarm probability: The probability that a target is detected when there is no target in the environment, expressed as $P_{\mathrm{FA}} = p(\hat{t} = 1 | t = 0)$. The probability of detection and false alarm follow an optimal trade-off according to the Neyman-Pearson lemma [53, Chap. 10]. The curve describing the detection probability as a function of the false alarm probability is called the receiver operating characteristic curve.

3) Mean squared error (MSE): The MSE describes the error in the estimation of target parameters (AoA, ToA, position, etc.). Assuming that the parameters to estimate from the received signal are $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_N]^\top$ and their estimates are $\hat{\boldsymbol{\theta}}$, the MSE is $\mathbb{E}[\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2]$, where the expected value is taken over realizations of the received signal.

4) Generalized optimal sub-pattern assignment (GOSPA) [54]: This metric combines the detection and

position estimation errors in one objective and it considers the association between the set of estimated targets and the true set of targets. The GOSPA objective is defined as follows. Let $\mathcal{P} = \{\boldsymbol{p}_1, ..., \boldsymbol{p}_{|\mathcal{P}|}\}$ and $\hat{\mathcal{P}} = \{\hat{\boldsymbol{p}}_1, ..., \hat{\boldsymbol{p}}_{|\hat{\mathcal{P}}|}\}$ be finite subsets corresponding to the true and estimated target positions, respectively, with $0 \leq |\mathcal{P}| \leq T_{\max}$, $0 \leq |\hat{\mathcal{P}}| \leq T_{\max}$. Let $d(\boldsymbol{p}, \hat{\boldsymbol{p}}) = \|\boldsymbol{p} - \hat{\boldsymbol{p}}\|_2$ be the Euclidean distance between true and estimated positions, and $d_\gamma(\boldsymbol{p}, \hat{\boldsymbol{p}}) = \min(d(\boldsymbol{p}, \hat{\boldsymbol{p}}), \gamma)$, where γ is the cut-off distance. Let Π_n be the set of all permutations of $\{1, ..., n\}$ for any $n \in \mathbb{N}$ and any element $\pi \in \Pi_n$ be a sequence $(\pi(1), ..., \pi(n))$. For $|\mathcal{P}| \leq |\hat{\mathcal{P}}|$, the GOSPA loss function is defined as

$$\mathcal{J}_p^{(\gamma,\mu)}(\mathcal{P}, \hat{\mathcal{P}}) = \left(\min_{\pi \in \Pi_{|\hat{\mathcal{P}}|}} \sum_{i=1}^{|\mathcal{P}|} d_\gamma(\boldsymbol{p}_i, \hat{\boldsymbol{p}}_{\pi(i)})^p \right.$$
$$\left. + \frac{\gamma^p}{\mu}(|\hat{\mathcal{P}}| - |\mathcal{P}|) \right)^{\frac{1}{p}}. \qquad (4)$$

If $|\mathcal{P}| > |\hat{\mathcal{P}}|$, $\mathcal{J}_p^{(\gamma,\mu)}(\mathcal{P}, \hat{\mathcal{P}}) = \mathcal{J}_p^{(\gamma,\mu)}(\hat{\mathcal{P}}, \mathcal{P})$. In (4), $\gamma > 0$, $0 < \mu \leq 2$ and $1 \leq p < \infty$. For larger p, a more severe penalty is applied to those estimates far from all ground-truth targets. The value of γ dictates the maximum allowable distance error. The role of μ, together with γ, is to control the detection penalization.

5) Cramér-Rao bound (CRB) [55]: The previous evaluation metrics empirically assess the performance of a sensing system, which are valuable to compare the performance of two different algorithms, but they do not convey any information about the optimality of an algorithm. The CRB gives a theoretical lower bound on the variance of an unbiased estimator and can be used for sensing applications. Assuming the received signal s_{Rx}, the variance of the estimated parameters $\hat{\boldsymbol{\theta}}$ is expressed according to the CRB as

$$\mathrm{var}(\hat{\theta}_i) \geq [\boldsymbol{I}^{-1}(\boldsymbol{\theta})]_{ii}, \qquad (5)$$

where $\boldsymbol{I}(\boldsymbol{\theta})$ is the Fisher information matrix defined as

$$[\boldsymbol{I}(\boldsymbol{\theta})]_{ij} = -\mathbb{E}\left[\frac{\partial^2 \log p(s_{\mathrm{Rx}}; \boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right]. \qquad (6)$$

In principle, all metrics described above can be used in the context of both spectrum and radar-like sensing. Detection and false alarm probability can be attached to the problems of finding the object's presence or occupation of a certain frequency band. MSE and CRB are used to characterize the localization algorithm's precision and achievable performance bound, and can be applied equally to object and emitter localization. Finally, GOSPA, a metric merging probability detection and localization precision, can be used in advanced scenarios, which involve either multi-target tracking or multi-source cognitive radio.

C. Machine Learning Basics

ML is a branch of AI that focuses on enabling computers to learn patterns from data and make predictions or decisions without being explicitly programmed [1]. Instead, a function $f(\boldsymbol{D}, \boldsymbol{\Theta})$ with input data $\boldsymbol{D}$ and parameters $\boldsymbol{\Theta}$ is used to approximate the prediction or decision task. The parameters $\boldsymbol{\Theta}$ of the function are then obtained using a dataset $\mathcal{D}$ with $\boldsymbol{D} \in \mathcal{D}$ and a performance measure $P(f(\boldsymbol{D}, \boldsymbol{\Theta}), T(\boldsymbol{D}))$, where the task to be learned is represented by $T(\boldsymbol{D})$. The performance

---

[3]We denote probabilities with $p(\cdot)$ as $P(\cdot)$ is used for the performance measure.

measure of numerous ML approaches is expressed as a *loss function*.

There are three primary learning paradigms: *Supervised learning* is based on a labeled dataset, where $T(\boldsymbol{D})$ is the known intended output, e.g., the ground truth of a parameter that $f(\boldsymbol{D}, \boldsymbol{\Theta})$ estimates. *Unsupervised or self-supervised learning (SSL)* finds patterns and structures in unlabeled data. In this case, $T(\boldsymbol{D})$ can include the variance of found patterns or different structures for clustering. *Reinforcement learning (RL)* teaches an agent to make decisions by rewarding or penalizing it based on its actions in an environment, potentially changing the environment with each action. The environment replaces $\mathcal{D}$, as each input $\boldsymbol{D}$ is dependent on previous actions taken by the agent. RL is of particular interest in the context of wireless networks, since it is researched as a candidate for complex decision-making in dynamic environments [56].

Focusing on supervised learning, random subsets $\mathcal{B}$ of $\mathcal{D}$ called *batches* (or mini-batches), with $|\mathcal{B}| \ll |\mathcal{D}|$ are usually propagated through $f(\boldsymbol{D}, \boldsymbol{\Theta})$. Then, the parameters $\boldsymbol{\Theta}$ are optimized iteratively using different $\mathcal{B}$. In each iteration, the loss function evaluates the mean performance over the current data batch. The MSE is a common choice for regression problems, i.e., tasks where a value dependent on $\boldsymbol{D}$ should be predicted such as estimating the delay from a received signal. For a batch $\mathcal{B}$, the MSE creates a criterion that measures the mean squared error between each element of the estimated value $f(\boldsymbol{D}_i, \boldsymbol{\Theta}) = \hat{x}_i(\boldsymbol{\Theta})$ and the corresponding ground truth value $T(\boldsymbol{D}_i) = x_i$ as follows:

$$\mathcal{L}_{\text{MSE}}(\boldsymbol{\Theta}) = \frac{1}{|\mathcal{B}|} \sum_{i=1}^{|\mathcal{B}|} (x_i - \hat{x}_i(\boldsymbol{\Theta}))^2. \tag{7}$$

The categorical cross-entropy loss is used for multiclass classification, i.e. $f(\boldsymbol{D}, \boldsymbol{\Theta})$ is tasked to differentiate between a number of different classes or types of samples in $\mathcal{D}$. Commonly, classification tasks are represented using one-hot notation, i.e., a sample of class c of C classes can be labeled using a vector $\boldsymbol{t}$ of size C with a 1 at the cth element and zeros otherwise. The categorical cross-entropy loss between the true one-hot labels $T(\boldsymbol{D}_i) = \boldsymbol{t}_i$ and the predicted distribution $f(\boldsymbol{D}_i, \boldsymbol{\Theta}) = \boldsymbol{p}_i(\Theta)$ is given by

$$\mathcal{L}_{\text{CrossEntropy}}(\boldsymbol{\Theta}) = -\frac{1}{|\mathcal{B}|} \sum_{i=1}^{|\mathcal{B}|} \boldsymbol{t}_i^\top \cdot \log \boldsymbol{p}_i(\boldsymbol{\Theta}). \tag{8}$$

For digital communication, the binary cross-entropy is often used for distinguishing between the binary states 0 and 1. This is a special case of the categorical cross-entropy with $C = 2$. It can be expressed as

$$\mathcal{L}_{\text{BCE}}(\boldsymbol{\Theta}) = -\frac{1}{|\mathcal{B}|} \sum_{i=1}^{|\mathcal{B}|} t_i \cdot \log p_i(\boldsymbol{\Theta}) + (1-t_i) \cdot \log(1-p_i(\boldsymbol{\Theta})). \tag{9}$$

After the loss function is calculated, the parameters $\boldsymbol{\Theta}$ are updated during training of $f(\boldsymbol{D}, \boldsymbol{\Theta})$. In supervised learning, an iterative gradient-based update is usually performed through backpropagation. By choosing different optimizers and optimizer parameters, the amount of change of $\boldsymbol{\Theta}$ is controlled in each optimization iteration. Then the next batch $\mathcal{B}$ is generated, allowing to iteratively update $\boldsymbol{\Theta}$ until a predetermined number

of iterations is reached or the performance measure reaches a target value.

There are multiple architectures in the literature that are used to implement $f(\boldsymbol{D}, \boldsymbol{\Theta})$. They often consist of so-called neurons that calculate a weighted sum of all inputs and add a bias. To enable approximation of non-linear functions, non-linear activation functions $\sigma(\cdot)$ are often applied to neuron outputs. Multiple neurons, whose outputs can be calculated in parallel, are grouped in a layer. In fully connected deep neural networks (DNNs), every neuron of one layer is connected to every neuron in the adjacent layers. Three other important neural network (NN) classes are convolutional neural networks (CNNs), recurrent neural networks (RNNs), and graph neural networks (GNNs). On the one hand, CNNs use convolutional layers that perform convolution of the input data with a trainable impulse response, denoted kernel in deep learning (DL) jargon, with parameters $\boldsymbol{\Theta}$. They are well suited to detect structures inside an input frame and therefore often utilized in computer vision tasks [57], including processing of spectrograms, range-Doppler maps, or constellation diagrams, where their shift-invariance is utilized. On the other hand, RNNs and transformer architectures are designed to capture (possibly long-term) temporal dependencies, and therefore suitable, e.g., for processing sequences of in-phase and quadrature (IQ) samples. The recently introduced transformer architecture is based on a self-attention mechanism, which relates different positions within a single sequence to compute a representation of the sequence [58]. Finally, GNNs are suitable for problems formulated as graphs. A graph is a mathematical representation of the relationships between different entities (called nodes) and their intrinsic attributes. GNNs can learn the attributes of the nodes, their relationships (called edges), or the full graph [59, 60]. In summary, however, there are no strict rules, and NN architectures allow a high degree of freedom to combine different layer types in application-specific architectures.

If the amount of training data is too limited or too specific, overfitting (or underfitting) can occur. In many ML domains like natural language processing or image processing, large text corpora or picture datasets are readily available. In contrast to that, wireless communications and wireless sensing research often relies on channel models instead of measured data. Depending on the ML task, obtaining models that approximate the real-world behavior well enough can be very challenging. A selection of stochastic geometry-based channel models has been standardized [61] by generalizing measured behaviors into a statistical model. Based on standardized models, channel simulators like QuaDRiGa [62] produce channel realizations, so that the performance of various algorithms can be evaluated and compared on these channels. Geometry-based stochastic channel models are less suitable for sensing applications, where deterministic quantities like target locations need to be estimated. In these cases, electromagnetic ray tracers like Sionna RT [63] or Remcom Wireless InSite [64] offer a reasonable compromise between the prohibitively high complexity of a finite element method (FEM) simulation of the electromagnetic field equations and oversimplified geometric models. Even with ray tracers, it is a good idea to specify and categorize particular environment scenarios to ensure

comparability, like in the DeepMIMO datasets [65]. There are also efforts to collect and publish datasets for various frequency ranges, for example, the NextG Channel Model Alliance by the National Institute of Standards and Technology (NIST) [66].

Remark. While ray tracing is sufficient when the dominant effects are geometric and deterministic, over-the-air data collection is required to replicate all real-world imperfections.[4] For example, real data is mandatory to validate the performance of the algorithms under real propagation, interference, and hardware impairments. Additionally, mobility effects or weather influences cannot be captured by deterministic models. Finally, empirical evidence is needed in the context of regulatory or standardization testing. In short, ray tracing can be seen as a tool for fast, scalable, and easily tunable modeling, while over-the-air data is required to prove the real life applicability of the designed algorithms.

Raw measurement data can often be represented in various domains and formats, such as time domain, frequency domain, beamspace, and sample autocorrelation. Digitized data is usually acquired as a complex-valued signal, which comprises the in-phase (I) and quadrature (Q) components taken at specific time intervals. These IQ samples are considered to be one of the richest representations of wireless signals for NNs that can be practically obtained. Conventional NNs with real-valued inputs and outputs remain common in wireless communications and sensing research. Usually, the real and imaginary parts of the complex-valued input vector are simply stacked and provided to the NN. Even though complex-valued neural networks (CVNNs) have been a research topic for a long time [68] and can natively process complex-valued signal representations, they have seen little adoption in the community, potentially due to limited support in common ML frameworks like PyTorch or TensorFlow at the time of writing.

Although NNs can adapt to arbitrary input domains, providing data in a reasonable format can significantly improve performance and reduce training time. Hence, feature engineering, i.e., converting the input into an appropriate format for processing by the NN, can lead to efficient ML solutions. For many sensing tasks, a common preprocessing step is to convert raw signal measurements and a known transmit signal into so-called channel state information (CSI) by performing channel estimation. Note that the dataset size can be reduced if only CSI instead of TX and RX signals needs to be saved, but this framework is limited to sensing scenarios where pilot signals are utilized for sensing or where a known transmit sequence is assumed at the sensing RX. The CSI is independent of the transmit signal and instead only depends on the wireless channel, which could potentially include the hardware effects of the TX and RX chains, as long as the transmission system is sufficiently linear. Since this is a useful signal representation for many sensing tasks, aforementioned channel datasets and channel simulators often directly contain or produce CSI instead of raw IQ samples. For communication problems, CSI can be used to simulate channel effects for particular transmit signals. On the downside, preprocessing input data

into features can also lead to loss of information. Models with higher capacity, e.g., those fed by multidimensional raw input data, can inherently learn the features of the wireless channel, including model deviations caused by hardware [6]. This approach often yields better performance at the cost of longer inference time and complexity.

There are multiple challenges in ISAC systems that seem uniquely suited to be addressed with ML. General models for channels and receivers exist, yet when components developed on classical channel models are integrated into a measurement setup, some loss of performance might be observed. This performance degradation can be due to model mismatches caused by hardware imperfections, distorted noise, quantization, or many other effects that have been ignored for the modeling of the channel. By training an ML system on real measurement data or even online using reinforcement learning, these effects can be addressed during optimization, even if only used for fine-tuning. Another point making ISAC specifically suited to approach using ML is that joint optimization of communication and sensing algorithms is often non-convex. It can be very hard to (approximately) solve these optimization problems, e.g., optimal resource allocation in ISAC, with classical approaches. It is still possible for ML solutions to get stuck at a local minimum, and one cannot guarantee that the found solution is a global minimum. Nevertheless, e.g., comparison with theoretical bounds known from estimation and information theory can help verify the convergence of ML systems.

D. Complexity Reduction

Efforts to reduce complexity include reduction of computational complexity, simplification of data acquisition, increased flexibility of algorithms, and increased accessibility of training methods. As ISAC research includes both spectrum and passive object sensing, we discuss it in a general section here.

The computational complexity of ISAC algorithms is an important factor for associated power consumption and hardware requirements, and can play a role in the algorithm's latency. Consequently, achieving similar performance while reducing complexity remains a vital direction to explore. For ML methods, we need to differentiate between the training and inference complexity. While training is usually performed once offline, being limited only by the time and hardware resources dedicated to training, the algorithm will be run many times with inference complexity on hardware with more limited resources. Research activities for complexity reduction tend to compare different ML solutions based on complexity that achieve similar performance. For trainable algorithms, the complexity of training and inference has to be considered separately. Complexity reduction is not unique to ISAC, but as ISAC algorithms are to run on cost-effective hardware with low latency, it is of special interest.

Model-based learning (MBL) [69] merges traditional model-based signal processing with DL by parameterizing traditional model-based signal processing to overcome the limitations of conventional methods. Assuming $g(\boldsymbol{D})$ is a function with no learnable parameters (e.g., a conventional model-based

---

[4]We note that accuracy of ray tracers can be improved via calibration by over-the-air collected measurements [67].

algorithm), it can be enhanced with learnable parameters $\boldsymbol{\Theta}$ to $\tilde{g}(\boldsymbol{D}, \boldsymbol{\Theta})$. How the classical and trainable components are entangled is up to the designer, often some tuning parameters of algorithms, such as weighting parameters, are made trainable. The function g and $\tilde{g}$ must be differentiable to enable backpropagation and learning of $\boldsymbol{\Theta}$. By incorporating domain knowledge, such as conventional receiver structures and algorithms, into the design of the NN architecture, this approach improves interpretability while improving training and inference efficiency and reducing the amount of training data required. Instead of training a DNN to perform a specific task from scratch, ML is employed to optimize the performance of existing algorithms, e.g., by tuning parameters to particular channel conditions and environments. A special type of MBL is deep unfolding, where g is an iterative algorithm, where each iteration is unrolled in order to finetune parameters $\boldsymbol{\Theta}$ for each iteration.

Methods like MBL and deep unfolding limit the amount of trainable parameters, while being able to achieve the same performance as pure ML solutions in some scenarios [69]. Choosing the algorithm structure and which parts of a classical algorithm benefit from being enhanced with trainable features is vital to achieve sleek and powerful solutions for complex problems. In [70], deep unfolding is used for bistatic joint sensing. This work optimizes the hyper-parameters of an ISAC algorithm for range-velocity estimation and communication demodulation, and then compares performance and complexity of this approach with other classical and trainable algorithms.

From a hardware-implementation point of view, complexity is a major indicator for cost of implementation and resource consumption. As an alternative ML paradigm, spiking NNs promise increased power efficiency in combination with neuromorphic hardware. In contrast to classical NNs, spiking NNs have emerged as a paradigm of neuromorphic computing promising higher energy efficiency [71]. Spiking NNs are designed to emulate the function of biological neurons by processing discrete spike events. As biological neurons are known to be very energy efficient, emulating them promises complex AI models with a fraction of the power consumption compared to current models. A neuron can have multiple inputs that contribute to the so-called membrane potential. When the membrane potential reaches a threshold, a spike is output by the neuron and the membrane potential is reset. Therefore, certain features such as clock-less computing, an inherent memory and a promise of high energy efficiency if implemented in hardware have attracted the attention of multiple research groups. In spiking NNs, different features of these neurons are made trainable. We refer the reader to [71] for a thorough introduction to spiking NNs. In [72], spiking NNs are developed to perform object detection and communication demodulation in an ISAC system. Although the paradigm shift to temporal spikes can be challenging, some applications such as impulse radio or pulse-based sensing can potentially be represented more easily in this signal space.

III. PASSIVE OBJECT SENSING

One entirely new capability which ISAC enables in future wireless communication networks will be the radar-like sensing
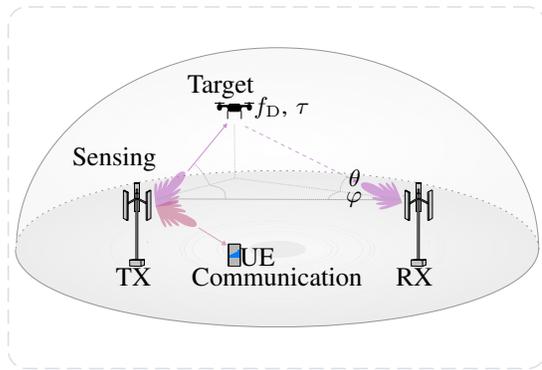

Fig. 3. A generalized ISAC setup for spatial sensing. Here, TX sends a signal to sense the target at RX and simultaneously communicates with the UE.

of passive objects and the capability to get environmental properties of the propagation environment. To this end, the wireless signal is exploited to identify reflections of radar sensing targets or to identify other environmental properties. Different flavors of ISAC radar sensing exist, ranging from transmitting dedicated sensing signals alongside communication signals to the full reuse of the available communication signals [13], [73].

The general sensing scenario for such an ISAC network is shown in Fig. 3. After capturing the reflected signals at one or multiple sensing receivers, the wireless channel is estimated and utilized in further signal processing steps to estimate targets or environmental properties. The foreseen deployments of ISAC in mono- and multistatic setups, the operation with standard communication hardware, and the constraint of economic feasibility create challenging tasks for the sensing signal processing. Alongside solving the actual sensing task with often limited spectral resources compared to dedicated radar systems, the design choices of the communication system impose additional constraints. Compared to the high integration of ISAC, conventional sensing solutions are employed in dedicated systems, where they are developed, tuned, and calibrated jointly with the dedicated measurement system to achieve their optimal operation points. Due to the seemingly uncountable and unforeseeable deployment scenarios of ISAC sensing setups, a full site-specific development is infeasible, and therefore ISAC mandates algorithmic solutions that are generalizable and computationally efficient.

In this section, we demonstrate how ML methods can be applied to ISAC systems to meet these challenges. Specifically, we highlight the trade-offs between sensing and communication, detection, estimation, tracking, and classification in ISAC radar sensing, as well as challenges in environmental and distributed sensing. Each part provides a brief introduction to the specific challenges, conventional solutions, issues, and presents novel ML solutions. We specifically explore the role of ML under the new paradigm shifts in ISAC and highlight the associated computational complexity and flexibility.

A. Communication-Sensing Trade-Off

The fundamental ISAC challenges arise from the combination of two historically separate systems. In this section, we consider the trade-offs unique to ISAC which arise either

from resource sharing or communication and sensing with the same TX signal. Fundamentally, ISAC systems are subject to both the deterministic trade-off and the subspace or power trade-off [74].

The deterministic trade-off refers to the distribution of the modulations symbols. On the one hand, communication systems are designed to maximize mutual information between two nodes. With sufficient signal-to-noise ratio (SNR), higher-order modulations are preferred, such as quadrature amplitude modulation (QAM). The transmit sequence needs to be random in order to carry information. On the other hand, sensing systems perform optimally with symbols with equal power on each subcarrier, such as quadrature phase shift keying (QPSK). This results in conflicting optimization goals that can vary in importance, as the desired quality of service for both systems can change over time.

The subspace or power trade-off refers to the transmit resources such as power or bandwidth being shared between sensing and communication tasks. In a general sense, it manifests as a resource allocation problem. In systems where time or frequency resource blocks are allocated to either communication or sensing, the main indicator of sensing performance is the amount of resources reserved for sensing. However, there is a performance gap between time/frequency sharing methods and known performance bounds [74], which is difficult to close with rigid resource sharing methods. ML can be used to explore trade-off spaces and find good operating points that meet design criteria for both functionalities or provide adaptive operating points. Additionally, it is of interest to shed light on the synergies of both functionalities and optimize the transmit signal. Conventional[5] optimization methods can be computationally expensive and inflexible, whereas ML can provide *good-enough* approximations, often at lower computational cost. In this context, RL has been shown as a suitable method to solve the complex resource allocation problem with lower computational complexity. For instance, in [75], a joint beamforming and power allocation problem is solved via RL to maximize the communication rate while ensuring sensing performance. In [76], power, bandwidth, and channel allocation are solved in a two-step RL framework. Other works also consider resource allocation with unmanned aerial vehicles, where the trajectory of the vehicles need to be planned to minimize flight time and avoid collisions [77], [78].

To illustrate the power trade-off, we use the example of beamforming. Depending on the sensing task at hand, different requirements on the beam patterns can be imposed by defining an area of interest for sensing. This area should be illuminated with a certain amount of power to enable detection of targets, follow the movement of a target, or scan different sectors consecutively. For communication, we can also define an area of interest as the area where communication RXs can be present. Different approaches aim to optimize the SNR at the communication RX, directly minimize the bit error rate (BER) or optimize a different indicator for communication performance. In this example, the beamforming weights b_ℓ

[5]Please note we use *conventional* solely to refer to existing non-ML solutions for simplicity.
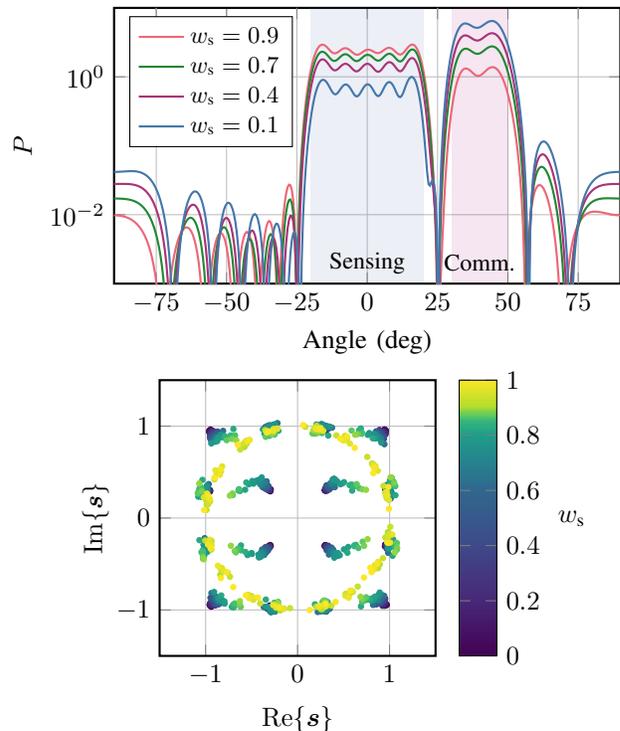

Fig. 4. Examples how power trade-off or deterministic tradeoff can manifest in an ISAC system from [79].

are the output of an NN and are optimized in an end-to-end ISAC system with the loss term

$$L(\boldsymbol{\Theta}) = (1 - w_\mathrm{s})\mathcal{L}_\mathrm{BCE}^\mathrm{commun}(\boldsymbol{\Theta}) + w_\mathrm{s}\left(\mathcal{L}_\mathrm{BCE}^\mathrm{detect}(\boldsymbol{\Theta}) + \mathcal{L}_\mathrm{MSE}^\mathrm{angle}(\boldsymbol{\Theta})\right),$$ (10)

with the NN parameters $\boldsymbol{\Theta}$, trade-off weight w_s and the loss terms referring to the definitions in (7)-(9) applied to the tasks of communication, detection and angle estimation. In Fig. 4, the resulting beam patterns are shown. The ISAC system trades off power between the area of interest for sensing and for communication. Depending on the importance of sensing indicated by w_s, power gets radiated either toward the sensing area or toward the communication RX.

Multiple research groups have addressed this trade-off using ML methods. In [80], [81], and [79], static beam patterns were obtained using a DNN in an end-to-end fashion, with the MSE of the AoA estimation, the cross entropy (CE) for target detection and the mutual information (MI) between TX and RX contributing to the loss term. In addition to common output normalization at the TX, legal output constraints of a wireless transmission are considered in [82] for beamforming in UE-centric sensing. The scenario in [82] consists of beamforming for sensing and communications with a choice of suitable base stations and generally leads to a non-convex optimization problem. A predictive beamforming approach is realized by a DNN, using the absolute error between the targeted side lobe level and effective isotropic radiated power in the loss function. Hybrid beamforming algorithms with multiple objectives are implemented in [83] using deep unfolding. A classical iterative optimizer with a fixed number of iterations is transformed into a DNN, enabling fast convergence. Transformers have also

been considered in predictive beamforming in [84], where a new transmission protocol is designed to reduce signal overhead and the complex beamforming optimization problem is solved via a transformer architecture. Other approaches consider that ISAC networks are composed of different devices, which can collaborate to provide optimal ISAC trade-offs. In this context, GNNs are suitable to learn the resource allocation of different devices in the network. For instance, in [85], vehicles are treated as agents in an RL framework, which can collaborate by sharing their observations to obtain a more efficient ISAC resource allocation. Moreover, in [86], the interference between different vehicles was also modeled and included in the GNN to maximize the communication sum-rate while maintaining a minimum sensing service. For instance, in [87], an ISAC transceiver communicates with users and senses targets in the environment. If the ISAC transceiver knows the number of users, a GNN can model the interference between the users and between the targets and the users to learn the optimal beamforming while maintaining a minimum communication service.

Concerning the deterministic trade-off, classical solutions to find good transmit signals have been hand-designed or limited to constant-modulus transmit signals to maximize sensing performance while enabling communication. Constellation shaping is being explored as a method to set flexible deterministic trade-offs, often being implemented using a DNN mapping the input bit to constellation symbols. Probabilistic constellation shaping, which uses QAM constellations but attributes different probabilities to different constellation points, is used to jointly optimize communication and sensing performance in [88]. Geometric constellation shaping, where the position of the constellation points is optimized, is illustrated in Fig. 4 for 16 constellation points. For high w_s, as defined in (10), the constellation tends towards a 16-PSK (sensing optimal) and with a trade-off tending toward communication with lower w_s it resembles a 16-QAM. Constellation shaping has been analyzed in more detail in [89] for orthogonal frequency-division multiplexing (OFDM)-ISAC systems. The authors showed analytically that constellation kurtosis is the deciding factor in possible sensing performance, independent of the used shaping method. The deterministic trade-off can therefore be flexibly used to improve either communication or sensing performance compared to classical methods. Therefore, a loss function optimizing the communication performance with an added constraint for the kurtosis can be used instead of an end-to-end evaluation of the communication and sensing performance, respectively.

B. Range-Doppler Detection

Target detection refers to the decision if or how many targets are present in a given area. It is commonly used the range-Doppler matrix (RDM), which describes the channels 2D power profile in range (measured as delay) and Doppler shift, exemplified in Fig. 6. The detection task is to identify which regions contain power reflected from a target, while keeping false alarms from clutter and noise to a minimum.

Generally, any waveform can yield an estimate of the RDM which is commonly used for estimating target parameters. We
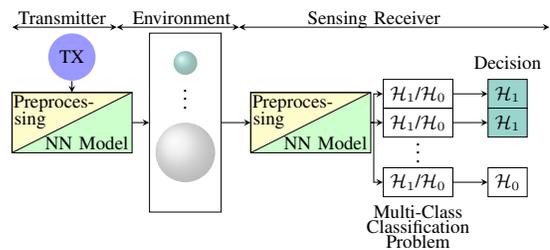

Fig. 5. Generalized framework for learning-based detection. The different targets (spheres in the environment) are represented by classes encoded on a grid, which is e.g., aligned with the bins of the RDM. Targets are detected by classifying each point in the grid as target or no-target. By extension, this approach is also suitable for multi-class classification of different target types.

show the calculation based on the OFDM waveform which is a main contender of becoming the ISAC waveform for 6G. The RDM is obtained from a series of M transmit $\boldsymbol{X} \in \mathbb{C}^{K \times M}$ and received $\boldsymbol{R} \in \mathbb{C}^{K \times M}$ OFDM symbols with K subcarriers after cyclic prefix removal. The M channel transfer function estimates are obtained by element-wise division of $\boldsymbol{R}$ and $\boldsymbol{X}$. Then the RDM is

$$\text{RDM} = |\boldsymbol{F}_K^* \left(\boldsymbol{R} \oslash \boldsymbol{X} \right) \boldsymbol{F}_M^\top|^2, \tag{11}$$

with $\boldsymbol{F}_K$ and $\boldsymbol{F}_M$ denoting the K- and M-point DFT unitary matrices, respectively. This is to say, we perform a column-wise IDFT to transform the frequency to delay domain, and row-wise DFT to transform the time to Doppler-shift domain.

A statistically optimal detector in AWGN is a Neyman-Pearson detector, which tests if a target is present at the channel estimate $\hat{h}$ at the discrete time delay n. It can be formulated as [53, Chap. 10]:

$$\frac{2}{\sigma_n^2}|\hat{h}_n|^2 \underset{\hat{T}=0}{\overset{\hat{T}=1}{\gtrless}} \chi_2^2(1 - P_f), \tag{12}$$

where $\chi_2^2(\cdot)$ denotes the chi-squared distribution with 2 degrees of freedom, P_f is the false alarm rate of detection and σ_n^2 denotes the noise power. While the Neyman-Pearson detector is statistically optimal, it requires an explicit estimate of the noise power, which is difficult to obtain in practice. Hence, more sophisticated scenarios with extended targets and clutter often utilize more practical constant false alarm rate (CFAR) approaches, such as cell-averaging- or ordered-statistics CFAR, which also provide a constant false alarm rate under AWGN noise. With CFAR the noise power for the decision threshold is estimated using training cells patterned around the cell under test, often accompanied by guard cells.

However, CFAR methods also observe performance impacts, when their statistical assumptions, e.g., about the noise, are not matched. For example, the estimate of the noise level is affected, when other targets are inside the training cells or the target extends over multiple cells. These also require additional filtering techniques, such as windowing, to avoid suboptimal performance. This not only causes increased false alarms but also prohibits the detection of closely-spaced targets. Other issues stem from false-detections caused by distortions from mismatched transmit signal structure, imperfect hardware calibration, clutter, or ghost targets. All of which are challenging to address and require additional specific filtering that is only applicable if the corresponding distortion source is known. Lastly, the distinct extended target shapes or

micro-Doppler can also cause multiple CFAR detections of the same target, resulting in additional estimation and tracking efforts. To combat those challenges, conventional algorithms require accurate modeling and subsequent, real-time applicable, filtering procedures to overcome these issues. Comparably, ML solutions may be suitable to overcome many of the presented challenges.

ML methods for detection can be generally described using Fig. 5. Each output of the model outputs is interpreted as a detector which is compared to a certain threshold during inference to decide if a target is detected. This allows general multi-target detection or joint detection and estimation if each class is associated with a certain estimation bin. ML components can be deployed both at the TX and at the sensing RX. For example, in [90] the waveform and detector are trained to approach a Neyman-Pearson detector, which is optimal in white noise. While no a priori information about the target and clutter model is assumed, their results show the trained solution achieves a more robust detection rate in clutter and colored noise compared to conventional approaches. Furthermore, [91] employs DL to successfully differentiate targets, such as vehicles, from clutter based on the RDM. It underlines how detection and object classification can be solved jointly, as a RDM may contain distinct target features which are lost in later processing stages. This is also underlined in [92], where the features of space debris targets are exploited for detection, via a combination of support vector machines and CSI-based sensing. Notably, the algorithm is trained based on simulated scenarios and measurement data to achieve accurate detection and classification of the debris type. In [93], the authors compare different CFAR versions with the CNN-method from [94] trained on simulation data. Their results highlight the ML models ability to separate closely-spaced targets in scenarios with similar magnitude, while otherwise performing similar to the CFAR methods. The discussion further emphasizes the additional need for further research to strengthen the understanding of the training procedures and simulation data models on the final model performance. In [81] considering multiple targets and [79] considering multiple snapshots, the comparison with a trainable detector and a classical statistical detector is performed. The simulations show it is possible to overcome the performance of generally optimal detectors without increasing false alarm rate, by using training data with scenario-specific features (e.g., derived from the area of interest).

C. Estimation

The estimation follows directly after a target is detected, with the goal to estimate the parameters of the propagation path caused by the target. Afterwards, the target can be localized by computing the possible locations with respect to the TX and RX positions. Therefore, the estimations' goal is to find the propagation path parameters of the target reflection with higher accuracy compared to the grid-based detections, to improve the overall sensing accuracy. Generally, the propagation path parameters relate a target's geometric position to its interaction with the impinging and reflected wave. The geometry of the
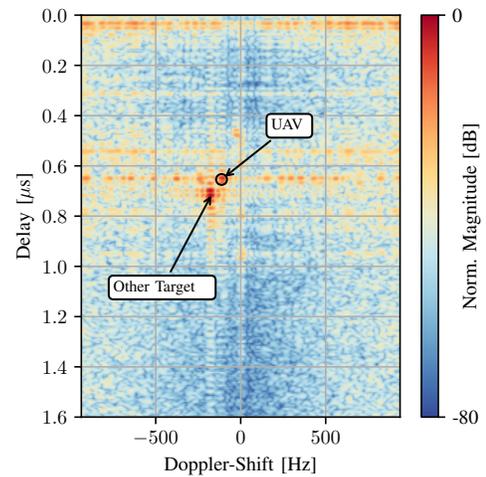

Fig. 6. RDM of an ISAC outdoor measurement from [95]. Strong clutter masks target visibility and limit target detection and estimation. A background subtraction removed clutter for better target visibility (UAV).
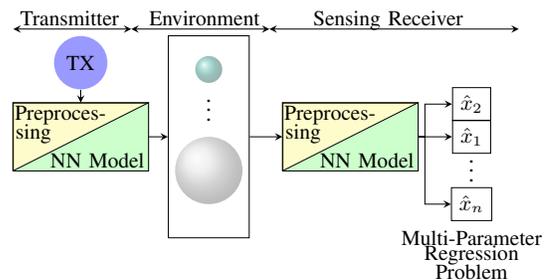

Fig. 7. Example framework for learning-based estimation. Compared to the multi-class classification task in Fig. 5, the model is trained for a regression task to enable continuous parameter estimates.

propagation is illustrated in Fig. 3. The reflection from a target in the signal model is described by its delay τ_ℓ, Doppler-shift $f_{D,\ell}$, and direction of arrival (DoA) in azimuth φ and elevation θ. While the DoA estimation typically requires antenna arrays at the RX (unless synthetic aperture radar (SAR) is used), both τ_ℓ and $f_{D,\ell}$ can be estimated from any single-input single-output (SISO) link. To model the wireless channel at the RX it is commonly assumed, that each path impinging on the RX is parametrized by its magnitude and phase, as well as τ_ℓ, $f_{D,\ell}$, φ and θ. However, due to the linearity of the channel, the measurement represents the superposition of all planar waves, posing a challenging inverse problem for estimation. For example, the paths can be insufficiently separated, even with strongly different magnitudes (strong masks weak), as the propagation paths in the channel depend solely on the propagation scenario.

Parameter estimation in ML usually relies either on joint detection as shown in Fig. 5 or on training a system on a regression problem as shown in Fig. 7. In the first option, the targets are estimated on a grid, limiting the resolution of the estimation to the grid spacing. Comparably, regression enables grid-less estimation, but requires separate estimation of the model-order, e.g., the number of targets, and a learnable parameter permutation/order. Both approaches may also be combined, similar to well-known computer-vision models [96],

and implemented in [97].

Conventional algorithms include simple axis-reading of the peaks detected in the RDM, and more advanced methods like multiple signal classification (MUSIC), estimation of signal parameter via rotational invariance techniques (ESPRIT), compressive sensing (CS), and maximum likelihood estimation (MLE) approaches. ISAC-suitable solutions must be real-time capable, provide sufficient accuracy and target separability under the spectral resource constraints of mobile communication networks, and handle the plethora of different hardware measurement setups, for which synchronization and calibration are only possible within the limits of the communication system. For example, iterative MLE estimators like RIMAX [98] achieve the limits of the CRB, but at a high runtime and significant computational cost, which is challenging to reduce [99]. Other challenges for conventional methods are imperfections arising from non-ideal measurement hardware effects, such as non-linearities, quantization, and imperfect calibration. Inevitably, such mismatches degrade the performance of algorithms reliant on the idealized signal models. Extended targets present another issue, as they appear as a superposition of insufficiently separated paths (in terms of sampling resolution) to an estimator based on (2). As a consequence, these paths create an additional overhead in the estimation, and possibly leveraging them for classification (as possible with ML, see III-B) requires additional hand-designed filtering.

As noted before, many flavors of the estimation task share the same mathematical structure, and as a result, publications in the area stem from many different scientific communities and often contribute to domain specific solutions. Therefore, we decided to split the presentation of the ML approaches accordingly and present them in the context of the specific estimation domain. We pick AoA estimation in a single dimension to introduce the mathematical background of harmonic retrieval problems, and then show the advances in different multidimensional estimation fields, focusing on the contributions of ML. We note that it is generally beneficial to perform a joint detection and estimation, as this always provides the optimal separation simply from more available dimensions for the separation, and hence detection, of targets. For example, a joint detection in range-Doppler-azimuth-elevation will yield better results compared to a range-only estimation. However, this is rarely possible under given hardware and real-time sensing constraints, as large chunks of data must be processed jointly, creating a need for lower-complexity algorithmic approximations. We therefore also emphasize these ML contributions where applicable.

1) Angle of Arrival: As described in (3), a waveform impinging on an antenna may yield a phase difference in the signal captured by different RXs. This phase difference is related to the angle θ from which the waveform comes. The problem of estimating θ from the received signal is commonly known as AoA estimation. Assume an array antenna estimating the AoA from M targets in the environment. Consider that the targets are in the far-field region of the antenna and that each target emits a signal $s_m(t)$ from an angle $\theta_m, m = 0, \ldots, M - 1$. Starting from the model in (3),

consider (*i*) that the targets have a LoS path with the antenna array, i.e., $L(t) = 1$ and (*ii*) the delay from each target to the antenna array can be estimated and compensated for, i.e., $\tau_0(t) = 0$. Then, the channel model for each target is

$$h_i(t, \tau) = \alpha(t)\delta(t - \tau_{i,0}(t))e^{-j2\pi f_c \tau_{i,0}(t)}. \quad (13)$$

Here, $\tau_{i,0}(t) = id\sin(\theta(t))/c$, $i = 0, \ldots, K - 1$, where d is the distance between RXs, c is the speed of light, and K is the number of sensors. In the case of a narrowband signal $s(t - \tau_{i,0}) \approx s(t)$.[6] Following (1) and considering the superposition of signals from different targets, the output signal at the i-th RX is

$$y_i(t) = \sum_{m=0}^{M-1} \alpha_m(t)s_m(t)e^{-j2\pi i \frac{d}{\lambda}\sin(\theta_m(t))} + n(t). \quad (14)$$

The continuous-time received signal in (14) is sampled and T transmissions (often called *snapshots*) can be collected for further processing. Aggregating the received samples across antenna elements and snapshots, the discrete-time received signal $\boldsymbol{Y}$ can be expressed as:

$$\boldsymbol{Y} = \boldsymbol{A}(\theta)\boldsymbol{S} + \boldsymbol{N}, \quad (15)$$

where $\boldsymbol{A} = [\boldsymbol{a}(\theta_1), \ldots, \boldsymbol{a}(\theta_M)] \in \mathbb{C}^{K \times M}$ is a matrix containing steering vectors of the form

$$\boldsymbol{a}(\theta) = [1, e^{-j2\pi d\sin(\theta)/\lambda}, \ldots, e^{-j2\pi d(K-1)\sin(\theta)/\lambda}]. \quad (16)$$

The sampled received signal for each target and snapshot are aggregated in $\boldsymbol{S} = [\boldsymbol{s}(1), \ldots, \boldsymbol{s}(T)]$, where the complex channel gain α_m is absorbed in s_m and $\boldsymbol{N}$ represents the additive noise.

There are different kinds of AoA estimation methods: (*i*) MLE methods, (*ii*) subspace methods, and (*iii*) CS methods. MLE methods attempt to solve the following optimization problem:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \mathfrak{L}(\boldsymbol{\theta}|\boldsymbol{Y}), \quad (17)$$

where $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_M]^\top$. The function $\mathfrak{L}(\boldsymbol{\theta}|\boldsymbol{Y})$ is the probability function of AoAs given the received signal, called the *likelihood function*. The likelihood function is in most cases multimodal, which hinders the implementation of MLE methods [100]. Several approaches have been proposed to solve this problem [100], [101], [102], but they require high SNR, large computational costs, or a small number of targets to estimate.

However, a common assumption of all AoA estimation methods is that the model of (15) holds. Under hardware impairments, methods degrade in performance due to model mismatch. In this context, ML can be harnessed to reduce the complexity of CS methods or to improve the estimation of subspace methods under few snapshots, coherent sources, low SNRs and modeling mismatches. The first step to AoA estimation with ML is to feed the received signal $\boldsymbol{Y}$ as input to a NN and obtain the AoA, i.e., following the notation of Section II-C, $\hat{\theta} = f(\boldsymbol{Y}; \boldsymbol{\Theta})$. To improve the angular resolution of the estimated AoA, the field-of-view of the receive antenna is discretized in angular candidates $\{\theta_i\}_{i=1}^N$ and the output of the NN is a probability vector that represents the probability that each angular candidate is the true angle, i.e., $\hat{\theta} = \sum_{i=1}^K \theta_i \hat{\boldsymbol{p}}_i$, with $\hat{\boldsymbol{p}} = f(\boldsymbol{Y}; \boldsymbol{\Theta})$. The AoA estimation

---

[6]Extensions for wideband signals have been proposed, but we skip their description for simplicity.

problem is regarded as a multi-class classification problem and the categorical CE loss is used to train the NN [103]. An extension of this approach is to discretize a smaller angular sector instead of the entire field-of-view, which provides a more accurate angular estimation [104]. A similar way to consider a discrete angular grid as output to the NN is to compute the spatial spectrum, defined as $|\boldsymbol{a}^T(\theta)\hat{\boldsymbol{p}}|$ for different values of θ. The estimated spatial spectrum computed from $\hat{\boldsymbol{p}}$ can be compared with the true spatial spectrum from the true AoA [105]. Other ML methods are inspired by subspace methods, which use an estimated covariance matrix of the received signal $\boldsymbol{R}_Y = \boldsymbol{Y}\boldsymbol{Y}^H/T$ to perform AoA. By definition, $\boldsymbol{R}_Y$ is Hermitian. The upper triangular part of $\boldsymbol{R}_Y$ contains all the information about $\boldsymbol{R}_Y$ and it can be used as input to a NN instead. Another approach is to use a vectorized and normalized version of the covariance matrix [106], i.e.,

$$\boldsymbol{r} = [\mathcal{R}\{\boldsymbol{R}_{1,2}\}, \mathcal{I}\{\boldsymbol{R}_{1,2}\}, \ldots, \mathcal{R}\{\boldsymbol{R}_{N,N}\}, \mathcal{I}\{\boldsymbol{R}_{N,N}\}]. \quad (18)$$

The input to a MLP can be $\boldsymbol{v} = \boldsymbol{r}/\|\boldsymbol{r}\|^2$ [107]. Moreover, the vector $\boldsymbol{v}$ can be reshaped back to a matrix and use a CNN to process the covariance matrix. The AoA can be directly computed as the output of the CNN or estimated with a probability vector based on a grid of candidate AoAs [107, 108].

Most limitations of subspace methods result from the mismatch between the true and estimated covariance matrix $\boldsymbol{R}_Y$. To improve the estimate of $\boldsymbol{R}_Y$, a CNN is trained to use an initial estimate of the covariance matrix (or its upper-triangular part) to produce a more accurate estimate, suitable root-MUSIC AoA estimation [109]. Another approach is to compute the MUSIC spectrum and then a MLP yields the estimated AoA [110]. Instead of training an estimation model, another direction in MBL is to learn efficient calibration procedures instead. In particular, for the case of hardware impairments, learning the parameters subject to impairments allows to calibrate the system. For instance, considering gain and position impairments in the ULA, the impaired steering vector becomes $[\tilde{\boldsymbol{a}}(\theta)]_k = g_k \exp(\jmath 2\pi p_k \sin(\theta))$, where g_k and p_k are the unknown gains and positions of the elements of the ULA. Considering as learnable parameters $\boldsymbol{\Theta} = \{g_k, p_k\}_{k=1}^K$ and using a subspace method that enables backpropagation, the performance of the impaired system can be improved [111]. As a summary, in Table III, we collect the most recent AoA estimation works, highlighting the problems from classical AoA estimation that they address.

2) Range-Doppler: Estimating the range and Doppler-shift of the detected targets enables the determination of a target's range and relative radial velocity. It is tightly coupled with the target detection (see III-B), as both operate on the same data domain. Typical algorithms include using peak positions from the RDM, subspace methods like ESPRIT, CS, and MLE approaches. The computational complexity of MLE methods is addressed in both [112] and [113], which present approaches to alleviate it through a combination of conventional MLE methods with ML. They illustrate that NNs can effectively initialize gradient-based MLE methods, reducing runtime while maintaining accuracy. These hybrid approaches combine the efficiency of data-driven methods with

the precision of model-based optimization, circumventing the need for exhaustive iterative searches.

Recent publications introduce ML-based solutions, particulary such that are trained on model-based data MBL instead of measurement data. Here the challenge is to design and train the model in a simulation environment such that it is suitable for the application in for actual real-world sensing applications. For example, [114] demonstrates it is possible to train a CNN for estimation on synthetic data and transfer it to measurement data, alleviating the need for costly measurement data collection. It is noted, that models like the presented CNN could also be re-used for target classification based on the RDM features similar to their use in computer vision applications. Traditional target classification typically involves a small set of classes [115], [116], and hence, computer vision models receives interest due to their successful classification with many classes. However, it must be noted that radio sensing measurements such as RDM provide less features for classification compared to camera recordings. Yet, similar models are successfully applied to classify targets from radar signals, involving three classes in [117], four classes in [118], and six classes in [119].

3) Angle-Range: The position of a target can be estimated transmitting a signal that contains different frequency components and using the same signal for different antenna elements. At the RX side, an *angle-range* matrix can be computed from the received signal across antenna elements and frequencies similarly to the RDM of (11). ML has been used in angle-range estimation to compensate for hardware impairments. Particularly, MBL was harnessed in [120] to compensate for hardware impairments while performing angle-range estimation and target detection. The considered impairments are inter-antenna spacing impairments, similarly to $\tilde{\boldsymbol{a}}(\theta)$ and the learnable parameters are the positions of the antenna elements in the uniform linear array (ULA). In Fig. 8, sensing results are represented as a function of the false alarm rate for the multi-target scenario of [120]. The MBL approach is compared with the baseline with perfect knowledge of the impairments and with a model-based calibration technique that requires no gradient backpropagation. Fig. 8 suggests that with MBL the performance of the system without impairments can be closely approached and with MBL the performance of model-based calibration can be outperformed. Unsupervised learning was introduced in a similar scenario in [121], where the received signal was processed to generate pseudo-labeled data to compensate for the impairments.

4) Angle-Range-Doppler: Angle-Range-Doppler estimation is a fundamental part of advanced driving assistance systems (ADAS). Light detection and ranging (LIDAR) is commonly used in ADAS since it is lightweight, cost-effective, and precise. However, the operation of LIDAR is severely affected by adverse weather conditions. Conversely, mmWave radar is more robust under harsh conditions at the expense of sparser clouds of points to detect targets. To improve the performance of mmWave radar for radar estimation, in [122] the accuracy of existing standard processing methods was improved with DL. Another use case of ML in radar estimation is to improve the resilience of the estimation against jamming, demonstrated

TABLE III
ML-BASED AoA ESTIMATION WORKS.

	Ref.	ML method	Low SNR	Few Snapshots	Complexity reduction	Hardware impairments	Coherent sources	Broad-band signals
DL	[103]	CNN	✓	✓				
	[104]	CNN		✓				
	[105]	CNN			✓	✓		
	[106]	Ensemble DL			✓			
	[107]	Denoising autoencoder				✓		
	[108]	CNN	✓					
MBL	[109]	Data-augmented root-MUSIC	✓	✓	✓	✓	✓	✓
	[110]	Data-augmented MUSIC	✓	✓	✓	✓	✓	✓
	[111]	Data-augmented MUSIC				✓		
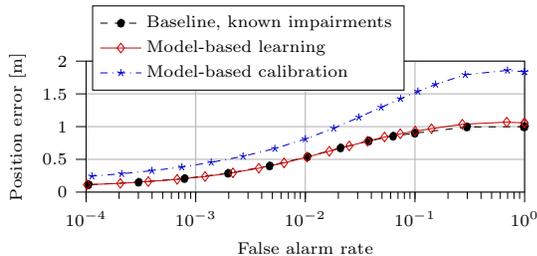
Fig. 8. Position error as a function of the false alarm rate. The results are computed from synthesized data considering a MIMO-OFDM radar with 64 antenna elements and 256 subcarriers. A maximum number of 5 targets in the scene are assumed, considering a random number in each snapshot. Inter-antenna spacing impairments were considered in the ISAC transceiver.

in [123], where ML is utilized to improve radar estimation against clutter and jamming. A scheme for a joint 4D detection and estimation of targets is presented in [124], based on MBL with data simulated for a uniform rectangular array (URA). There the CNN approach from [113] is extended to jointly estimate the delay, Doppler-shift, azimuth, and elevation parameters of an unknown number of paths using simulation data. Their results show, that joint processing of 4D radar sensing data can enable significant runtime advantages and high accuracy when combined with MLE methods.

D. Tracking

So far, we have considered that we obtain an estimate of the target position and velocity every time that one or multiple snapshots were transmitted. These estimates over time form the trajectory of each target in the scene. If target trajectories are considered instead of independent estimates, the estimate of the target position and velocity in the next time step can be improved. Tracking refers to the estimation of these trajectories. The major challenge of tracking is to associate new measurements with object trajectories including: (*i*) creating and removing trajectories and (*ii*) distinguishing clutter from targets. There have been several algorithms proposed for multi-object tracking, such as the probability hypothesis density [125], the generalized labeled multi-Bernoulli [126], and the multi-scan trajectory Poisson multi-Bernoulli mixture [127]. However, the unknown measurement-target association makes these approaches have

a super-exponential complexity in the number of time steps to process [128]. ML can reduce the computational complexity of conventional tracking algorithms. At the TX side, tracking is used to predict the user position and perform beamforming, which can help in the design of narrower beams with larger antenna arrays to increase the data rate of users. Using no labeled data at the TX side was studied in [129] for predictive and adaptive beamforming. An optimization problem with complicated constraints was transformed in an unconstrained problem, whose objective was used as pseudo-labels to train the DL model. Predictive and adaptive beamforming was studied in [129], where an unsupervised approach with an long short-term memory (LSTM) network is proposed. The work in [130] also uses an LSTM network with the raw data instead of a two-step approach where the positions of targets are estimated and then beamforming is performed. Strategies to reduce the training overhead have also been proposed [131]. At the RX, given the temporal dependencies of the observations in tracking tasks, transformers are a suitable solution to estimate the number of targets in the environment and their trajectories with a suitable computational complexity [132]. This solution was also applied to improve the performance of existing approaches for trajectory smoothing, where all noisy trajectory samples are available and the goal is to retrieve most likely trajectory given the noisy past observations [133]. MBL solutions have also been developed based on the Kalman filter (KF). The KF provides the miminum mean squared error (MMSE) estimator for a time-varying discrete system under AWGN. It is based on linear equations that describe how the system evolves over time, referred to as the state-space model. However, many problems involve non-linear state-space models. Several solutions have been proposed, such as the extended KF [134] and the unscented KF [135]. The family of particle filters [136] was also introduced for non-linear, non-Gaussian state-space models, which obtains better results than the KF with higher computational complexity. The main weakness of the aforementioned filters is that they rely on the state-space model and filters exhibit performance degradation if the state-space model is not accurate. Several works deal with some level of uncertainty in the state-space model, but they cannot achieve the same performance as the filters with full

knowledge. MBL methods can enhance conventional filters to adapt to unexpected deviations from the assumed state-space models. Regarding the KF, the work in [137] augments the KF to deal with non-linearities and model mismatches. This approach was extended to make the NN robust against changes in the model [138] or to quantify the uncertainty in the estimations by extracting the error covariance matrix [139]. The particle filter was also augmented with DL to mitigate the model mismatch [140].

E. Gesture Recognition

Apart from the propagation parameters discussed in previous sections, targets also possess other useful features, depending on the sensing application. For example, we already discussed applications for target classification, such as pedestrian, bike, car, from the radar sensing data. Here, we focus on gesture recognition, particularly with multi-sensor fusion techniques.

Gesture recognition is one of the most important aspects of human-computer interaction, with applications such as the development of hearing aids, monitoring of a patient's emotional state, or lie detection [141]. At first, gesture recognition was based on visual aids [142] or wearable sensors [143]. However, these sensors exhibit issues under inappropriate light conditions and usability issues. In contrast, the use of radar signals is robust against light conditions and users do not need to carry sensors. While model-based algorithms rely on wave propagation theory and statistical models, requiring hand-crafted feature extraction [144], ML can combine feature extraction and classification and be trained to jointly optimize both tasks. Gesture recognition under the lens of AI was first considered using synthetic data. The work in [144] performs simulations with publicly available datasets. In [145], the radar echoes are first transformed into range-Doppler time points, which are further processed by DL. Recently, ML has also been successfully deployed in practical implementations. In [146], mmWave access points were used to perform gesture recognition and in [147], the authors use a 4D imaging radar —a type of radar that can measure elevation, azimuth, range, and Doppler of targets— to classify between different human postures.

F. Environmental Sensing

Environmental sensing refers to sensing applications, where the material parameters or a particle distribution in the air is measured instead of estimating geometrical features. An early example is radar for weather forecasting which has been in commercial deployment since the 1950s. With different available bandwidths and DTs becoming more accessible, additional environmental features become available for sensing.

A more recent example for weather prediction is given in [149], which uses fused radar images and wind forecasts to classify between different weather classes with an NN. In this work, the open MeteoNet dataset was used for training with a CE loss function to achieve more accurate predictions compared to conventional optical flow benchmarks. In [150], the rain fade for satellite communications is predicted using
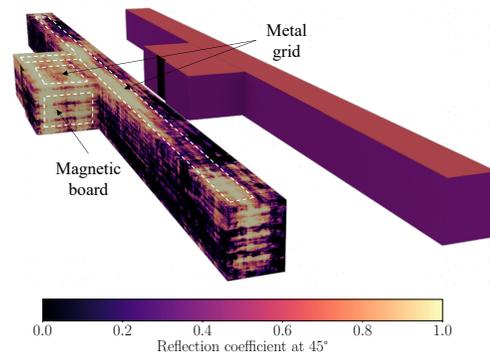

Fig. 9. Learned materials from [148] to obtain fine our rough representation of the environment. This graphic is under ⓒ ①CC-BY 4.0 International license.

DL from satellite radar imagery, as well as link power measurements. The authors utilize a CNN to predict the probability of rain fade from the satellite radar images. In the community of weather prediction, data fusion is generally well established as multiple different sensor types are available, including radars. In the future, this set of sensors can be complemented by the ISAC system to enhance weather prediction services.

In the field of material parameter sensing, [151] gives an overview of ML algorithms for THz sensing to measure the parameters of solid and gaseous materials. Lastly, in [148] parameter sensing based on a known geometry is proposed as a way to improve the accuracy of virtual twins. Through ray tracing, channels accurately representing the actual environment are simulated. It is still unclear how much detail is needed to build the environment and materials. In Fig. 9, a hallway is shown with the measured reflectivity of the walls, floor, and ceiling. Depending on the complexity of the targeted material model, the radiation environment can be approximated with few materials or on a fine grid.

G. Distributed Sensing

Distributed sensing is generally of interest, as 6G networks are expected to provide more sensing nodes, including mobile nodes, such as UEs. Different concepts for distributed sensing are discussed, such as a network composed of multistatic links [13] and more advanced multi-sensor MIMO systems with multi-sensor coordination [152]. Regardless, preprocessing the initial measurements of the wireless channel at the nodes, e.g., via target detection and estimation, is essential to avoid significant signaling overhead. Therefore, we consider distributed sensing also as a fusion task, where estimates from multiple nodes must be jointly processed. The problem of distributed sensing is therefore twofold. Firstly, sensing data representation must be communicated without too much overhead, and secondly, a fusion algorithm must be capable of estimating the localization results. Such an algorithm must handle fusion of different links, frequency bands, and possibly even additional sensor types, such as cameras or LIDARs. For such concepts, multimodal fusion based on ML is proposed in [22]. In [153], the feature and confidence extraction for different modalities benefit from ML, but the data fusion itself consists of classical fusion based on the confidences

of the different sensors. This approach seems promising from a signaling overhead perspective. On the other hand, joint estimation and fusion through ML for more difficult scenarios [154] could boost sensing performance. A thorough overview of multimodal fusion of ISAC can be found in [155]. Although classical fusion is prevalent in ISAC, ML is a common tool in the data fusion literature [156] and might show successful applications when distributed multimodal sensing tasks receive more attention.

H. Datasets

Most datasets for object sensing provide channel measurements instead of transmit and receive signals to reduce the dataset size. While this limits the ability to study low-level waveform design or end-to-end signal processing, this constraint can often be circumvented in practice by applying any TX waveform on top of the available channel measurements. For many high-level tasks, the provided channel representations remain sufficient for training and evaluating algorithms. TX side ML is often based on simulated channels using channel simulators such as Sionna [30]. Datasets based on measurement data for radar sensing include [95] which is suitable for distributed sensing or the collection of DeepSense datasets [157] which support multi-modal sensing and many different application scenarios. The DISC dataset [158] provides measurements of indoor channels obtained at 60 GHz for detection, tracking or activity recognition. Yet, simulated datasets such as Sensiverse [159] are also prevalent in ISAC research, as the effort in obtaining a measurement dataset with the associated tagging of target positions is high. While simulated data can be very helpful for prototyping ML solutions or comparing different algorithms, real measurement data is necessary to characterize tasks in more detail.

I. Outlook and Lessons Learned

ML research for ISAC is flourishing, with numerous ML solutions being proposed for different sensing tasks. However, similar tasks are required for diverse applications, leading to variations in the scenario assumptions and rendering direct comparisons difficult. Currently, comparisons mostly rely on classical benchmark algorithms, as no universal dataset for algorithm comparison is available. Most proposed ML algorithms for sensing solve a specific sub-task within the sensing signal processing chain. Since pre- and postprocessing are critical, model-based learning has emerged as a valuable technique. Looking ahead, we expect the adoption of many specialized AI algorithms rather than a single unified solution, which aligns well with the distribution of network functions in a communication system. We elaborate on this vision in Section V.

IV. Spectrum and Emitter Analysis

As 6G networks are supposed to be more intelligent and context-aware, passive radio sensing technologies are gaining prominence. These methods allow network devices and infrastructure to obtain environmental and spectral insights in an opportunistic manner, i.e., without active transmission, enabling a wide range of applications. This section addresses sensing and localization tasks performed by sensors without an illumination source. More specifically, we focus on ML-powered methods for dealing with non-cooperative TXs and aim to answer the question of who is transmitting from which location, which is crucial, for example, for spectrum management, resource allocation, and jammer detection. Often overlooked in the context of ISAC, these tasks are essential in CR and cognitive radar, and their broader implementation could greatly improve the intelligence and resilience of future wireless networks.

A. Spectrum Sensing

Spectrum sensing aims to detect underutilized bands in the monitored portion of the spectrum. The concepts of spectrum sensing and spectrum sharing underpin ideas of CR [160] and cognitive radar [161]. Specifically, Listen-Before-Talk (LBT) function is a part of many medium access control (MAC) protocols in wireless communications. LBT is a key mechanism for the sharing of spectrum in an uncoordinated way, detecting primary signals and spectrum opportunities [162]. The current trend toward the coexistence of various wireless technologies in the same spectrum imposes even stricter requirements for their coordination [163]. Traditional detectors like energy detectors, matched filters, and cyclic feature detectors achieve good performance under ideal conditions where the assumed models of channels and signals are accurate and known in advance. However, this assumption is often unrealistic in practical CR systems, resulting in problems caused by model mismatches that can degrade performance. To overcome these issues, the potential of AI-enhanced methods for spectrum sensing in the context of CR became evident early on [164], followed by reviews at the beginning of the 2010s [32], [33].

We begin by formulating the problem of wideband spectrum sensing, which is crucial for both current and future wireless networks as it enables multi-band opportunistic channel access. Single-band spectrum sensing can be viewed as a special case of this. In wideband sensing, an RX monitors the spectrum that consists of N_f frequency sub-bands. The signal $\mathbf{s}_{RX}$ received at the RX front-end is expressed as shown in (1). Subsequently, to extract the spectrum features of the received signal, power spectrum density (PSD) is computed by applying the Fourier transform to the autocorrelation function, i.e.,

$$\mathbf{S}_{RX_n} = \text{FFT}[\text{Corr}(\mathbf{s}_{RX_n})], \qquad (19)$$

where $\text{FFT}[\cdot]$ denotes the fast Fourier transform, and $\text{Corr}(\cdot)$ is the signal autocorrelation. Note that since the time domain signal from (1) contains noise, the spectrum $\mathbf{S}_{RX_n}$ is also noisy. The sub-band measurements are then concatenated to represent the full band spectrum $\mathbf{S}_{RX}$ as

$$\mathbf{S}_{RX} = [\mathbf{S}_{RX_1}, \mathbf{S}_{RX_2}, ..., \mathbf{S}_{RX_n}, ..., \mathbf{S}_{RX_{N_f}}].$$

The detector's task is to decide on each sub-band's n occupancy based on noisy measurement vector $\mathbf{S}_{RX}$. The decision represents a binary hypothesis testing problem formulated as
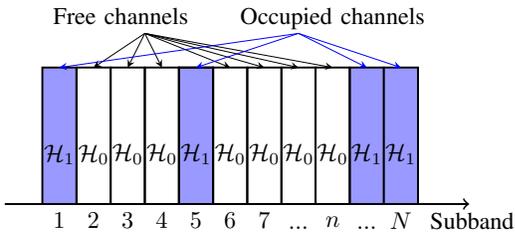
Fig. 10. An illustration of a multi-band spectrum sensing. Free and occupied sub-channels are marked with '$\mathcal{H}_0$' and '$\mathcal{H}_1$' labels, respectively.
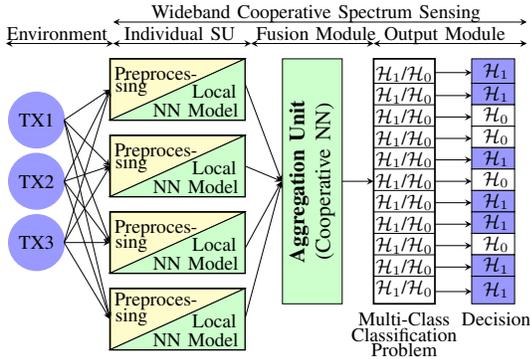

Fig. 11. Generalized architecture for learning-based wideband cooperative spectrum sensing.

$$f(\mathbf{S}_{\text{RX}_n}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \tag{20}$$

where $f(\cdot)$ is some function of measurement $\mathbf{S}_{\text{RX}_n}$, $\mathcal{H}_0$ and $\mathcal{H}_1$ denote the absence and presence of (primary) signals, respectively, and γ is the decision threshold. The value of γ is associated with the detection probability, with the default choice being 0.5. In practice, it has to be tuned to balance the importance of missed detection and false alarm events for a particular application. Fig. 13 shows the wideband spectrum sensing scenario, where only a few sub-bands are occupied.

Over the recent years, a huge number of research articles applying ML-based methods to the problem of spectrum sensing were published. Next, we overview methods based on a single- and multiple-node spectrum observation.

1) Single-User Spectrum Sensing: Single-node spectrum sensing is based on independent judgment of a single user. It allows for an easier system that does not require data fusion at the cost of limited performance. In [165], authors presented a CNN-based method capable of adapting to various noise types and detecting the presence of various signals, including those unseen. They used the signal power spectrum as CNN input and the maximum-minimum eigenvalue ratio-based method as a baseline [166]. The results achieved under various false alarm probabilities show that the NN-based approach performed better than the baseline, especially for low SNR. The *DeepSweep*, another CNN-based model that exploits real-world recorded IQ samples of Wi-Fi signals as input, was presented in [167]. Focusing on narrowband interference detection in the 2.4 GHz IMS band via subsequent processing of the spectrum parts, the authors report a real-time inference capability of less than 1 ms with up to 98% accuracy. In [168],

a spectrum detector is implemented based on a Transformer architecture. The main idea of this work is to exploit the self-attention mechanism to learn not only the inner-band but also the inter-band spectral features in a wideband scenario. To benchmark their model, the authors use an earlier CNN-based framework called *DeepSense* from [169], reporting a significant complexity reduction and higher detection accuracy. In practical wireless systems, obtaining a sufficient amount of training data is often challenging and not realistic. This issue was addressed in [170], where the authors generated additional labeled data using a GAN framework. The authors of more recent [171] followed this concept of data augmentation with a GAN to train their CNN-based sensing algorithm with improved generalizability and robustness.

2) Cooperative Spectrum Sensing: Cooperative spectrum sensing refers to an approach in which a decision on spectrum occupancy is based on multiple observations of the area of interest. Exploiting cooperative sensing nodes distributed in the area of interest promises increased coverage, performance, and robustness [172]. For instance, collaborative spectrum sensing mitigates the hidden node problem in CR networks based on carrier-sense multiple access with collision avoidance [173]. Furthermore, distributed observations enable the design of more complex and flexible data fusion pipelines and learning paradigms, resulting in diverse research challenges.

In [174], a hierarchical method that integrates convolutional layers with LSTMs was proposed. This study models sensing units (SUs) equipped with multiple antennas that receive signals over various time slots, presenting spatial features as correlation matrices that are processed through the CNN and LSTM layers. Subsequently, features extracted separately for each RX are concatenated and passed through additional LSTM and fully connected layers, followed by a Softmax function, which estimates the probabilities of $\mathcal{H}_0$ and $\mathcal{H}_1$ and decides on channel occupancy. The authors also provided calculations of γ based on a targeted false alarm rate. The reported inference time is 1.3 ms with simulations performed on NVIDIA Quadro P5000 GPU 16GB RAM, thus indicating the real-time feasibility of the algorithm. The authors of [175] presented an unsupervised collaborative DL method based on PCA and dimensionality reduction, obtaining performance comparable to supervised learning. The PCA stage performs representation learning from unlabeled data, while the dimensionality-reduction stage provides pseudo-labels that enable discrimination between channel occupancy states. A recent work [176] focuses on wideband cooperative spectrum sensing under partial observations, i.e., a scenario where each of RXs monitors only a few consecutive sub-channels N_f, whereas the TX can occupy any sub-channel. The original CNN for a multi-task problem is decoupled into sub-band relevant sub-networks, reducing the computational cost without performance impairment.

3) Sensing-Throughput Trade-off: Unlike radar, spectrum sensing is purely passive and opportunistic, i.e., there are no transmit sensing signals, and no fundamental communication-sensing trade-off similar to those discussed in Sec. III-A for radar-communication systems (i.e., at the waveform or signal design level). However, there is a resource trade-off between
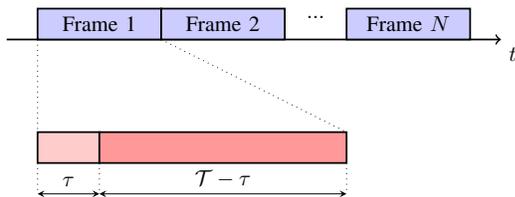
Fig. 12. Sensing-throughput trade-off illustration: A schematic frame structure for cognitive radio networks with periodic spectrum sensing. Sensing and data transmission slot durations are denoted as τ and $\mathcal{T} - \tau$, respectively.
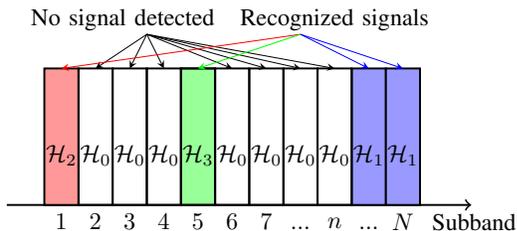

Fig. 13. An illustration of a multi-band multi-signal detection and type classification. $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{H}_3$ denote 3 signal types, while $\mathcal{H}_0$ denotes the signal absence.

how long or often one can sense and how much one can transmit, which is called the sensing-throughput trade-off in CR networks [177]. Longer sensing time improves detection reliability but reduces the time available for data transmission. Fig. 12 illustrates a generic frame structure designed for a CR network with periodic spectrum sensing, where each frame consists of one sensing slot and one data transmission slot.

The sensing-throughput trade-off can be improved or even eliminated by using dedicated sensing nodes that analyze spectrum continuously while communication nodes transmit, thus decoupling sensing and communication in time. Furthermore, using dedicated sensing nodes enables full potential of cooperative spectrum sensing. However, some trade-offs and challenges remain:

- Resource trade-off: To share their results, sensing nodes require a backhaul or control channel. This channel can be wired or wireless.
- Extra nodes consume power, which might be relevant in the context of power-constrained IoT networks.
- In the case of cooperative sensing, the collected data needs to be transmitted and processed, which adds delay.

These trade-offs are not easily tractable, thus motivating their data-driven optimization via ML approaches.

B. Signal Classification

Non-cooperative signal classification can be seen as an advanced version of standard spectrum sensing and is sometimes referred to as intelligent spectrum sensing [38]. The main difference is that it requires not only to decide if the subband is free or not but also to recognize the signal type. Fig. 14 illustrates the generic signal classification problem.

For the input x, which can be raw data or feature vector, signal classification can be represented as the multiple hypothesis testing problem:

$$\hat{\mathcal{H}} = \underset{n \in [1, |\mathcal{S}|]}{\arg\max} f(\mathcal{H}_n | x, \Theta), \tag{21}$$
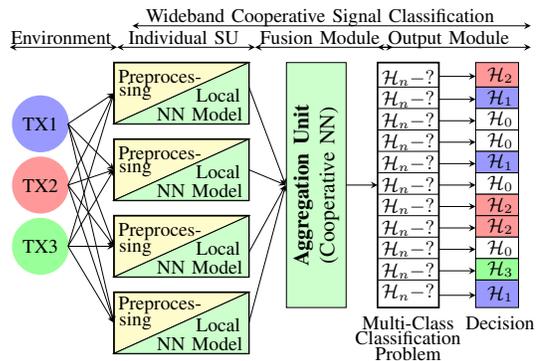

Fig. 14. Generalized architecture for learning-based wideband cooperative signal classification for 3 TX signal types.

where $\mathcal{H}$ and $\hat{\mathcal{H}}$ are the ground-truth signal type and the predicted signal type, respectively; $\mathcal{S}$ is the pool of possible signal types with $|\mathcal{S}|$ being the number of signal types in $\mathcal{S}$; $f(\Theta)$ is the mapping function from samples to signal types, where Θ denotes the model parameters.

Signal classification can be roughly arranged into two categories: (*i*) a broader task of waveform classification (or radio access technology (RAT) recognition), and (*ii*) AMC. By classifying waveforms, networks can verify the legitimacy of signals. For instance, detecting anomalous or unauthorized waveforms could signal potential security threats like jamming or spoofing [39]. This helps improve the security and integrity of communication systems. AMC is a more specific task that can enable real-time adjustments to RX settings and accelerate the demodulation procedure. The interest in AMC can be traced back to the turn of the 1990s, with [178] and [179] exemplifying early works on analog and digitally modulated communication signals classification, respectively. Nowadays, wireless networks use numerous modulation schemes, protocols, and waveforms, making signal classification way more complex. Moreover, wireless signals can overlap or undergo distortion due to interference or fading, challenging the recognition task.

The methods developed over the next decade can be categorized into two categories which are based on the likelihood function and statistical features of the signal, respectively [180]. Since feature-based approaches are more practical and easier to implement, they attracted significant interest. The list of popular statistical features includes a ratio of in-phase and quadrature power components, the standard deviation of the direct instantaneous phase, the standard deviation of the absolute value of the normalized instantaneous signal amplitude, the mean value of the signal magnitude, etc. [181] A popular set of statistical features can be illustrated by higher order moments (HOMs), defined for a complex-valued input sequence x of any length as

$$M_{pq} = \mathbb{E}[x^{(p-q)} \cdot (x^*)^q]. \tag{22}$$

Furthermore, combinations of HOMs can be leveraged to obtain higher order cumulants (HOCs) denoted as C; for a full list of formulas, refer to [181]).

Classification based on a single feature is hardly possible due to the abundance of signals from various wireless standards

that exhibit similar statistical parameters. In principle, all extracted features can be taken into account for signal classification problems, serving as input to the ML-based classifier. However, the number of features used to train the algorithm affects the computational complexity and inference time. Hence, for practical applications, the proper feature selection is of great importance to choose the most distinguishable and discard those that are redundant. To this end, methods from statistics were adopted to quantify the dissimilarity between features. For instance, Bhattacharyya distance (BD) is often used in feature selection tasks to compare feature probability distributions based on their overlap and shared information [182]. For two discrete probability distributions P and Q in the same domain $\mathcal{X}$, BD is computed as

$$D_B(P, Q) = -\ln \left(\sum_{x \in \mathcal{X}} \sqrt{P(x)Q(x)} \right). \tag{23}$$

The BD is always non-negative: a value of 0 means that the two distributions are identical, while higher values indicates a greater difference between them. The task of effective feature selection based on BD in the context of AMC was studied and evaluated on three neural network-based classifiers for AWGN and frequency-selective fading channels in [183].

The rest of the subsection provides an overview of popular ML-based signal classification approaches, datasets, and the evolution of the signal classification task with advancements in wireless technology.

1) Knowledge-Driven Signal Classification: Vectors of statistical features, such as moments, cumulants, and cyclic cumulants of the signal itself, were used as input in early ML-based signal classification applications [184]. These features can be fed into simple decision trees [185], random forests [186], or SVMs [187]. This approach remains popular nowadays since it results in lightweight ML models and promises robust performance under poor SNR conditions. For instance, in 2016, the authors of [181] mapped 20000 signal samples into 21 statistical features as an input to a small DNN to classify 7 modulation types, namely BPSK, QPSK, 8-PSK, 16-QAM, 64-QAM, 2-FSK, and 4-FSK. Fig. 15(a) shows the standard deviation of the absolute normalized centered instantaneous frequency for the signal segment. Based on this feature alone, it is already possible to uniquely separate 3 classes. Furthermore, the 4 rest classes have a rather small overlap. In contrast, [188] studies waveform classification for potential B5G scenarios, considering a synthetic dataset comprised of various realizations of 8 waveforms, including 3GPP-compliant 4G, 5G, Wi-Fi, and Bluetooth waveforms for communications, linear frequency modulated waveform (LFMW) and pulsed waveforms for radar application, and two fully integrated waveform candidates based on OFDM and phase-modulated continuous waveform (PMCW). Fig. 15(b) shows normalized histograms which approximate pdfs of the underlying distribution of a cumulant $C_{21} = \mathbb{E}[|x|^2]$, where x is an input time series. Due to significant feature overlaps, it is difficult to expect an accurate signal classification in this scenario. This example shows that methods driven solely by
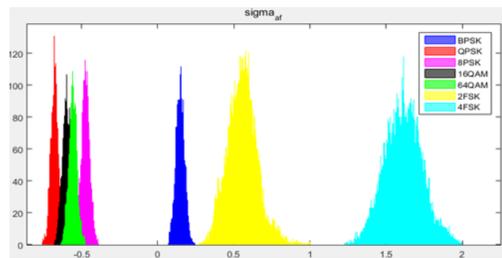
(a) Standard deviation of the absolute normalized centered instantaneous frequency for the signal segment for AMC [181](©[2016]IEEE)
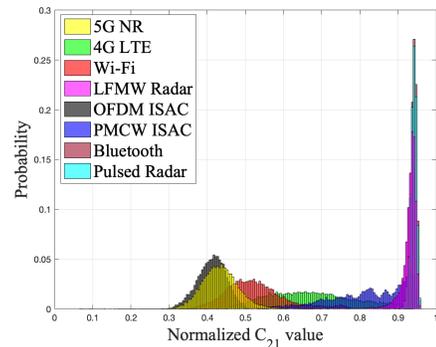
(b) Normalized histogram illustrating the probability distribution of statistical feature C_{21} of waveforms in the dataset from [188].

Fig. 15. Histograms that illustrate the probability distribution of statistical features of signals in the datasets.

expert knowledge, which operate on a feature space of reduced dimension, have limited potential for certain tasks.

2) Data-Driven Classification with Raw Input Data: To circumvent the aforementioned limitations of the feature-based approach, T. O'Shea and J. Hoydis, in their seminal article from 2017, proposed utilizing raw time-series data as NN input [6]. Using raw data, namely IQ samples or their amplitude-phase (AP) representation, increases computational complexity but promises enhanced classification accuracy. The classifier architecture in [6] itself is a straightforward CNN composed of narrowing convolutional layers followed by FC layers and a dense Softmax output layer. Following that work, many researchers studied various aspects of improving this purely data-driven learning approach. For example, the authors of [189] employed a similar CNN-based method and concentrated on inference time reduction, directly targeting real-time capability. In [190], the signals were classified into in-network users, out-network users, and jammers coexisting in the same wireless network. Similarly, this paper used IQ data as input to a CNN. A CNN-based AMC that accounts for phase imperfections is performed in [191], where inputs are represented by images of modulation constellations with dimensions of 256×256, each containing 1000 points. However, since the original raw time-domain IQ samples represent a 1D time series, architectures based on standard RNNs, LSTMs, and Transformers are of special interest and have gained popularity. Another study on AMC for OFDM systems leveraged various signal representations, specifically IQ and AP, as bi-stream input, thus utilizing the synergy between distinct signal modalities [192]. In this paper, the input

sequence length equals 2048. Additionally, an attention-based CNN-LSTM model has been used to design the classifier.

Although data-driven methods are currently far more popular due to their superior classification accuracy, they lack explainability. Combining raw data with the expert-extracted features and prior knowledge not only makes a step towards better explainability but also aims to reap the benefits of both approaches [193]. Finally, when exploiting raw data as an input for classification, the performance-complexity trade-off should be wisely considered. The signal should be adequately represented by a sufficient number of samples to enhance classification accuracy while avoiding excessive computational burden to maintain low inference time.

3) Signal Classification with Spectrograms as Input: Spectrograms are a popular alternative signal representation for classification tasks. A spectrogram visually represents the spectrum of frequencies in a signal as it changes over time. They are generated by dividing the original time-domain signal into overlapping segments and using the short-time Fourier transform (STFT) to determine the magnitude of the frequency spectrum for each segment [194].

The conversion of wireless signals into spectrograms allows the application of image processing techniques for signal classification. Specifically, an adapted version of You Only Look Once (YOLO), a supervised CNN-based algorithm to learn how to find and classify the objects in an image, was successfully leveraged for wideband signal recognition task in [195]. This work introduced a two-step lightweight AMC framework: In the first stage, YOLO was exploited to detect, localize, and classify narrowband signals in wideband spectrograms, and in the second stage, a custom CNN refined the classification results.

Since constructing spectrograms requires computationally heavy processing stages that allow some degree of freedom, it is vital to be aware of several involved parameters on the algorithm performance. In this context, [196] studied the influence of STFT window size and type on AMC accuracy for their designed CNN-based algorithm. The reported result, performed over 26 signal types affected by multipath fading channels, shows the marginal dependency of AMC accuracy on STFT with 128, 256, and 512 samples for window overlap of 25%, 50%, and 75%, and for six window functions. The study shows some advantage of using shorter window lengths and bigger overlaps. However, it lacks inference times for different methods, which is important for fair trade-off estimation.

Overall, despite their apparent merits, spectrograms may be impractical for real-time wireless signal classification applications. First, they require additional signal processing steps, which can be computationally expensive. Therefore, it is recommended to take into account preprocessing time while evaluating the inference of the ML-based algorithms that rely on spectrograms as input. Second, although spectrograms ensure a rich signal representation, they lack information regarding the signal phase [197].

4) Jamming and Anomaly Detection: Discovering jammers, intentional interferers, and other types of spectrum anomalies can be seen as a subproblem of signal classification, dealing exclusively with the unwanted TXs. Jamming is often studied for wireless applications in the industrial domain, where it can be a serious issue, as equipment downtime can cause multimillion-dollar losses. ML techniques are expected to play a crucial role in detecting jammers in wireless networks by identifying anomalies in environments caused by malicious TXs. For example, [198] studied an autoencoder-based jammer detection and mitigation for automated guided vehicles in mmWave networks for manufacturing environments. Further, [15] presents an ML-based framework for anomaly and jammer detection in the context of DTs for industrial environments. To detect spectrum anomalies, the authors employed unsupervised learning based on distributed received signal strength (RSS) measurements. Specifically, the algorithm compares legitimate users' expected radio environment map estimates with actual distributed RSS measurements, thus detecting the unauthorized TX presence and roughly localizing it.

In summary, since jammers often target critical infrastructure, AI-enhanced detection algorithms should be able to continuously sense the environment and detect attacks as soon as possible. Hence, the learning algorithms' real-time capability and adaptability are essential. For this purpose, unsupervised and RL techniques that allow the model to update with new data and maintain responsibility for new types of jamming attacks require detailed investigation. We refer readers to [39] for a comprehensive taxonomy of anomaly and jammer attacks along with state-of-the-art learning techniques to detect and mitigate them.

5) Cooperative Signal Classification: Similarly to cooperative spectrum sensing, cooperative signal classification by multiple RXs observing the same environment can extend the monitored coverage area, offer spatial diversity, and remove the dependency on a single node and channel which might be under poor conditions. Standard cooperative algorithms can be exemplified by [199], which performs likelihood ratio-based distributed fusion using multiple radios, where each RX makes a local classification decision and then these decisions are merged to make a global decision. An alternative approach is to fuse the data in a centralized way, where the data can be raw [200] or be represented as a set of statistical features, e.g., cumulants [201]. Note that when done not properly, the fusion of data from multiple RXs can even result in degraded performance compared to a single RX solution.

A prominent work [202] studies DL models for signal classification task in a distributed sensor network only using averaged magnitude FFT data. This approach resulted in a practical model which comprises i low-cost sensors which scan narrow bands at frequencies f_i from 50 MHz to 6 GHz, and transfer small amount of data to the cloud, where LSTM-based models perform classification on the over-the-air dataset which consists of 6 waveforms. The reported accuracy varies depending on a specific network structure, and yields more than 80% for the largest model. Further, the CNN was used in [203] in the context of multiple-input and multiple-output (MIMO) systems, where the RX, equipped with multiple antennas, classifies the modulation types. Signals received at each RX antenna are fed to CNN, which obtains sub-results, which are subsequently merged

to make the final decision. Four cooperative decision rules, namely direct voting, weighty voting, weighty averaging, and direct averaging are applied to infer the final modulation type, with weighty averaging reported to be superior. The performance of centralized and distributed CNN-based AMC were compared in [204]. The centralized approach performs slightly better, but this comes at the cost of higher training time and increased requirements for the edge computer's computational power. A recently proposed collaborative AMC framework for OFDM signals [205] focuses on signal-level fusion. As a result of multi-stage signal processing, a signal distribution property vector is constructed, which is finally fed to a simple lightweight NN which comprises 2 convolutional and 2 fully connected layers with multi-class entropy loss function.

Currently, there are relatively few papers that utilize distributed observations for various signal classification tasks using ML. However, this topic has significant potential, with various problems to tackle. Table IV summarizes the literature on spectrum sensing and signal classification.

6) Scenarios and Types of Signals to Recognize: For decades, the standard tasks were to classify modulation formats, waveforms or other parameters of pure communications signals. Recently, due to spectrum congestion and cognitive radar and radio evolution, the inclusion of radar for classification signals has become more common [209]. The advancement of radar-communications convergence will impose more difficult challenges on wireless RXs and networks in general to ensure correct operation.

Signal classification was recognized as an important task that could be tackled in the future ISAC RX design [35]. Learning of modulation schemes and/or radar sensing and communications signals waveforms can be used instead of, e.g., successive interference cancellation (SIC), thus reducing RX complexity and facilitating faster acquisition. A step in this direction is [210], where the authors presented a CNN-based waveform classifier for ISAC RX design which operates on time–frequency images. However, this work only considers a co-existence radar-communications scenario. In 6G and beyond, the inclusion of potential fully integrated ISAC signal waveforms is under investigation. For instance, in such a case, radar-centric PMCW or communications-centric OFDM can be added to the standard set of waveforms. Since the ISAC waveforms are the tweaked versions of the existing dedicated radar and communications waveforms, they retain a strong resemblance, thus hindering the classification. It can be envisioned that in such a scenario, only processing raw IQ data by a relatively large DNN might obtain acceptable classification accuracy.

In general, waveform and RAT classification are more difficult compared to AMC because they require analyzing the signal in a broader scope and considering a wider range of features: Protocols, modulation schemes, access techniques, and combinations of these parameters. For example, 4G LTE, 5G NR, and Wi-Fi waveforms are all OFDM-based and use similar sets of modulation schemes, therefore, achieving high classification accuracy can be very challenging. In contrast, AMC is more focused on identifying a particular characteristic (modulation scheme) of the signal, making it a less complex task overall, although still challenging under low SNR or with high-order modulations. Fig. 15 illustrates the difference in complexity of tasks by examples of statistical features distributions. In literature, AMC dominates the field; however, it can be anticipated that blind waveform and access technology recognition will play a significant role in future wireless networks, contributing to tasks such as interference monitoring and management, network security, and anomaly detection. In this context, ML techniques will be essential to improve automation and intelligent aspects. It can be safely envisioned that with the growth of modulation orders used in wireless systems, types of RATs, and waveforms, intelligent spectrum sensing in 6G networks will require novel learning approaches capable of balancing high performance with acceptable complexity for real-time operation.

7) Datasets: Common benchmark datasets are of high importance to compare various ML models for radio signal analysis. RadioML 2018.01A, a canonical dataset for emitter types classification used in hundreds of academic research articles since 2018, was generated in [211] and improved in [212]. It contains signals of 24 radio signal types generated and recorded under varying wireless propagation environment using software defined radio (SDR) platforms. However, as the authors themselves admit, those datasets have several known errors, and while remaining useful for academic and educational purposes, should not be used within real products.[7] Nevertheless, these works laid the foundation for a recent RML22 [213], where the authors claim a more realistic and correct methodology for signal generation. Apart from a dataset, its main contribution is the open source Python-based framework that allows the generation of wireless signals impaired by a variety of propagation effects, which can be used to evaluate ML models for modulation classification.

A wideband dataset that enables detection and localization in time and frequency domains is presented in [214]. This dataset is represented by synthetically generated data which emulates wideband signals through a combination of diverse layouts of narrowband emissions. The authors of [215] provide another wideband spectrogram images dataset containing 20000 labeled samples of Wi-Fi and Bluetooth signals along with tools to create new data sets for specific requirements.

While the aforementioned datasets consist exclusively of communications signals, modern wireless environments allow radar users to use the same spectrum. Given this, [216] provides a collection of pulse radar and 4G LTE signals recorded over the air in the shared Citizen Broadband Radio Service (CBRS) band (3.55-3.7 GHz), where signals coexist in overlapping and non-overlapping manner in a number of varying signal-to-noise-and-interference ratio (SINR) conditions. The authors also describe the data collection setup, consisting of three Ettus X310 USRP SDR systems, one each for the radar TX and 4G LTE signals in the CBRS band and a third USRP for receiving IQ samples.

Table V summarizes several important datasets for various radio sensing modalities.

---

[7]https://www.deepsig.ai/datasets/

TABLE IV
AN OVERVIEW OF THE ML-BASED APPROACHES FOR SPECTRUM SENSING AND SIGNAL CLASSIFICATION FROM LITERATURE.

Task	Ref. (Year)	Input	ML method	Distr.	Reported contributions	Limitations
Spectrum Sensing	[174] (2023)	Covariance matrices of energy correlations	Hybrid CNN-LSTM	✓	High accuracy and robustness under imperfect CSI	High computational complexity
	[167] (2024)	IQ samples	CNN-based		Real-time capability (<1 ms inference time), 98% accuracy	Limited functionality, only narrowband operation
	[168] (2024)	Rearranged PSD	Transformer		Operates in wide band, reduced complexity, higher detection accuracy	Requires validation with over-the-air measurements
	[176] (2024)	Rearranged PSD	CNN-based	✓	Wideband cooperative sensing solution under only partial observations	High inference time, not suitable for real-time 6G applications
	[169] (2021)	IQ samples	CNN-based		Over-the-air setup, real-time capability (0.61 ms inference time), 98% accuracy	Relatively low accuracy and high inference time with a large-scale dataset
	[170] (2018)	Samples in time domain (not detailed)	SVM and Random Forest		Demonstrates advantages of GANs in training data augmentation for sensing	Results are obtained in a basic narrowband scenario
	[175] (2022)	Energy vectors	Unsupervised learn. using Gaussian mixture model and PCA	✓	Useful for scenario with little labeled data, robust in low SNR	Noticeable performance gap compared to supervised learning
	[206] (2024)	Energy vectors	Attention-enhanced RL	✓	Reliable sensor scheduling, reduced communication overhead	Unclear real-time capability and practical feasibility, limited performance accuracy comparison
Signal Classification	[181] (2016)	Statistical feature vectors	Small DNN with 5 FC layers		High accuracy AMC	Limited and relatively simple set of low-order modulated signals with distinct statistical properties
	[6] (2017)	Raw IQ samples	CNN-based		AMC with higher accuracy compared to feature-based AMC	Brute force solution with high computational complexity
	[192] (2024)	Raw IQ samples and their AP form	Attention-based hybrid CNN-LSTM		High-accuracy AMC for OFDM systems with over-the-air captured data	Limited to OFDM waveform, too high inference time (>1 s)
	[196] (2022)	Spectrograms	CNN-based		Studies dependency of AMC accuracy on STFT with 128, 256, and 512 samples	No information regarding inference time
	[202] (2018)	Averaged FFT magnitudes	LSTM-based	✓	Experimentally proven low-cost efficient AMC approach	Limited accuracy (around 80%) even with a small number of classes
	[203] (2022)	Raw IQ samples	CNN-based	✓	Presents a discussion on four cooperative decision rules in AMC	Small number of candidate classes, low-order modulations (BPSK, QPSK, 8PSK, 16QAM), no information on complexity
	[205] (2024)	Signal distribution property vector	Lightweight NN based on CNN	✓	Effective AMC under unbalanced SNR and varying channel conditions	Only 6 low-order modulation formats, requires real-world validation, no discussion on complexity
	[207] (2020)	(i) IQ and (ii) FFT	CNN		RAT classification in sub-GHz implemented with low-cost SDRs and 2 NN input types	Only ultra narrowband signals, relatively high inference time
	[208] (2021)	Spectrograms	YOLO-based		Adoption of a popular ML object detector for efficient detection and localization of signals in wideband RF spectrum	Limited recognition accuracy of some signals, required real-data validation
	[209] (2020)	Fourier synchro-squeezing transformation (FSST)	CNN-based		Studies waveform classification in radar-communications coexistence scenarios	Show good performance only with high SNR; no information regarding inference time

Overall, pioneering RadioML 2018.01A from 2018 remains the most popular dataset for testing wireless signal classification approaches. While it is still useful to examine the performance of the designed algorithms, this dataset does not represent the realms of wireless standards of 2026. More specifically, we now have moved from the middle of 4G to 5G-Advanced, with even early experimental 6G systems available for research, which resulted in evolved signal standards (new waveforms, much higher order modulation schemes), hardware characteristics (base stations, user equipment, and SDR platforms with different characteristics), wireless environments (due to ubiquitous IoT, introduction of new frequency bands), and overall higher performance requirements. Therefore, to embrace these developments, the research community apparently needs to adopt a new benchmark; however, it is difficult to dictate or predict when and what it will be. The vast majority of articles on ML-based signal classification evaluate models whose input is a signal captured by a single RX. A distributed spectrum analysis approach, where networked sensors monitor the same environment, could be of interest for signal classification as well as jamming and anomaly detection. Further, its potential advantages and trade-offs (classification/detection accuracy vs. complexity) can be investigated. In this context, novel datasets emulating distributed sensing scenarios are welcome. In principle, modern software tools allow fast standard-compliant signal generation using MATLAB Communication Toolbox, while the realistic propagation environment can be modeled leveraging ray tracing software like Altair Winprop or Sionna RT. Further, over-the-air real signals can be recorded for various scenarios using SDRs (for more details on wireless data generation and collection, see Section II-C).

C. Machine Learning-based Remote Transmitter Positioning

Obtaining low-complexity and good-performing positioning algorithms is often challenging due to the nonlinear relationship between the signal source coordinates and positioning parameters, i.e., as time-difference-of-arrival (TDoA), AoA and RSS. Indoor environments likely present non-line of sight (NLoS) propagation conditions that significantly damage the performance of model-based position estimators. Moreover, traditional positioning parameter modeling neglects hardware-related impairments, such as power amplifier nonlinearity. Dealing with such issues usually means designing computationally intensive approaches. In such a context, ML-based techniques are able to circumvent the nonlinear characteristics and provide a lower complexity solution.

TABLE V
OVERVIEW OF SOME DATASETS USED IN RADAR AND SPECTRUM SENSING RESEARCH. OTA DENOTES OVER-THE-AIR MEASUREMENTS.

Dataset / Reference	Paradigm	Signal Rep.	Main task / Application	Scale	Source	Key characteristics and limitations
RadioML 2018.01A [212]	Spectrum	Raw IQ	AMC	2M samples (21 Gb), 24 modul. classes	Synthetic	Legacy benchmark, some technical flaws were discovered
RML22 [213]	Spectrum	Raw IQ	AMC	10 classes, 420.000 samples, 128 pairs IQ	Synthetic	Improved RML22; generation code available; limited scenarios
[214]	Spectrum	Spectrograms	Signal recognition	550.000 wideband signal samples	Synthetic	Frames of multiple overlapping signals
[215]	Spectrum	Spectrograms	RF frame detection	20.000 samples, 512x512 pixel images, mainly Wi-Fi and Bluetooth	Synthetic	Generation code available; limited number of standards
[216]	Spectrum	Spectrograms	Signal detection	55.5 Gb of LTE and radar samples in 3.5 GHz range	OTA	Real RF signals; limited diversity
ISAC UAV Dataset [95]	Radar, ISAC	IQ samples	Emitter and target localization	Data is captured at 3.75 GHz; 100 MHz bandwidth	OTA	Real RF signals; very specific scenario
DeepSense [157]	RF, vision, LiDAR, radar	Diverse	Multi-modal sensing incl. beyond radio	More than 1 million data points	OTA	Foundational dataset; limited flexibility
DISC [158]	ISAC	CIRs	Human activities detection and recognition	Mm-wave (60 GHz) CIRs; hundreds GB in total	OTA	Dataset for human sensing; specific conditions
Sensiverse [159]	ISAC	CIRs	3D reconstruction, moving target detection	CIRs for TX-RX pairs over 4 square km scenes; very large scale (15 TB)	Synthetic	Extra large, simulated, primarily for vision-based ISAC research

Just like in radar-like sensing, the performance of the designed TX positioning algorithms can be benchmarked by CRB [217] (for details, see Sec. II-B). Note, however, that CRB is a physics-informed bound on the estimation error given the underlying measurement model (e.g., based on TDoA, AoA, RSS and CSI statistics). In the context of ML-based methods, it is sometimes used to tell whether the model performs close to physical limits, since ML methods ultimately rely on the same noisy input data [218]. However, since ML algorithms usually introduce bias (e.g., due to regularization) and are limited by finite training data, one should be aware that a classical CRB should not be interpreted as the true minimum achievable error for ML algorithms.

Related works on ML-based localization seek to represent the positioning parameters, such as RSS, CSI, time-of-arrival (ToA) or AoA in image-like formats to exploit the proven capabilities of CNNs architectures used in image processing applications [219], [220], [221].

1) Blind Transmitter Positioning: Remote positioning assumes that within an area of interest, multiple SUs are connected to a central unit (CU). The SUs receive the incoming signals from the signal source(s) and send them to the CU, where its coordinates are estimated. The source coordinates can be communicated to the source, or be only employed for network management. Positioning can be further classified as cooperative or non-cooperative depending on whether the devices aid in their position estimation, e.g., by transmitting positioning-specific signaling. Private wireless networks that are threatened by jamming transmissions can benefit from non-cooperative remote positioning for reliability.

Within an area of interest, a total of N SUs are placed in fixed locations and communicate their measurements to a CU. Blind TX positioning refers to a remote positioning approach where the sources that need to be localized do not aid in the process. In this case, position estimation can be obtained through AoA or RSS measurements. In the AoA case, SUs have to be equipped with an antenna array for estimating AoA, consequently increasing the SUs size and cost. On the other hand, RSS measurements can be obtained through simpler and low-cost SUs.

Large-scale propagation models are used to predict the RSS based on the distance between the source and SUs. These models typically combine deterministic and random components and are applicable for distances beyond the Fresnel distance, corresponding to far-field propagation conditions. The deterministic component follows an exponential decay, where the decay exponent varies depending on the environment. Some models also incorporate an additional attenuation factor to account for fixed obstructions, such as walls, glass windows, and floors. Meanwhile, the statistical component captures power fluctuations caused by multipath propagation and random blockages.

The RSS at the j-th SU can be expressed by the lognormal path loss model (PLM) given by

$$P_j = P_0 - 10\beta \log_{10}(d_j/d_0) + n_j, \qquad (24)$$

where β represents the decay exponent, n_j is Gaussian distributed shadowing noise, d_j denotes the Euclidean distance between the source and the j-th SU, and P_0 is the power received at the reference distance d_0, which should be larger than the Fresnel distance.

The lognormal PLM aligns relatively well with empirical data, but is mainly used for coverage prediction. Thus, it is not a positioning-specific propagation model. Unfortunately, this model does not lead to a low-complexity closed-form position estimator, which means that approximations must be applied to obtain real-time solutions. Hence, ML is an attractive solution to RSS-based positioning. Recent investigations have shown that a relatively small fully-connected NN outperforms model-based positioning techniques [222].

2) Coherent Cooperative Transmitter Positioning: In positioning systems, all relevant parameters are estimated from the received signals. Theoretically, utilizing these raw signals as

input to an NN should enable positioning accuracy comparable to classical methods, as they inherently contain all the necessary information about the source's location. However, a significant challenge arises when training such a model as the number of required training examples increases exponentially with the input dimension, thus rendering the learning process computationally expensive and difficult to generalize effectively. To address this issue, one common approach is to reduce the dimensionality of the received signals by extracting specific features, such as the RSS. By summarizing the signal information into a more compact representation, the computational complexity is reduced, allowing for more efficient model training. However, this simplification comes at a cost, as valuable information about the characteristics of the signal is inevitably lost in the process.

In this context, feature engineering plays a critical role in ensuring that ML algorithms perform optimally. The success of a model is directly influenced by how well the input data is represented, as different representations can significantly impact learning efficiency and accuracy. The primary objective of feature engineering is to derive representations that capture the underlying structure of the data in a way that facilitates learning and generalization. By carefully selecting and transforming relevant features, it is possible to enhance model performance while maintaining computational efficiency. Striking the right balance between dimensionality reduction and information retention is key to developing robust and scalable positioning solutions.

If the SUs are clock synchronized, the CU can coherently exploit the position-related information in the received signals, and ML offers us the flexibility for engineering an enhanced position information parameter. In [223], the authors define a position information correlation matrix (PICM) where the main diagonal matrix contains the RSS values at each SU, and the off-diagonal elements of the matrix indicate the correlation levels between sensing units, which depend on their relative positions with respect to the source. By analyzing both the diagonal and off-diagonal elements, it is possible to gain insights into the spatial distribution of received signals, improving the accuracy of ML models.

D. Multi-Task Learning for Spectrum Analysis

The information extracted from wireless signals via spectrum monitoring can be used for various tasks. It is, however, possible to go further and leverage the same input to solve multiple tasks jointly, thus yielding extra functionality. The main argument behind the multi-task learning (MTL) concept is that such an extra functionality can often be obtained at a reasonable DNN complexity cost and, if the tasks are related, it can even improve the performance of one or a few tasks [224].

For instance, detection of the unused or underused portions of a radio frequency spectrum can be done jointly or subsequently with signal classification. This straightforward MTL application is explored in [225], where the authors propose an SSL approach to spectrum sensing in which a DNN is trained without any labeled data or prior knowledge of the transmitting signals. Specifically, the captured radio

signals, represented as one-dimensional FFT sequences, are processed by a CNN-based deep clustering architecture that jointly learns feature representations and groups similar signals into clusters. These clusters, whose number is not predefined, are later assigned semantic labels such as "4G LTE", "5G NR", "Wi-Fi", or "noise" through a simple manual mapping step. By adopting this SSL-based deep clustering framework, the model autonomously learns discriminative spectrum features from unlabeled data. When evaluated on three real-world datasets and compared against autoencoder-based state-of-the-art architectures, the proposed approach achieves superior clustering performance, while also requiring fewer parameters and lower computational complexity.

Furthermore, analysis of the same signal can be used to extract the information related to the TX position. Based on this premise, the authors of [226] proposed a framework to simultaneously classify wireless signal waveforms and locate the TX. To this end, they model an indoor scenario with a single TX, which broadcasts a signal captured by several distributed sensors in the form of IQ samples. Subsequently, portions of these IQ samples are used as input for a dual-task DNN, which jointly estimates the wireless signal waveform classes and the TX coordinates. The framework is refined, extended, and further analyzed in [227]. The loss function comprises weighted cross-entropy loss and MSE for waveform classification and TX localization, defined by (7) and (8), respectively. While the values of $\mathcal{L}_{\text{CrossEntropy}}$ and $\mathcal{L}_{\text{MSE}}$ characterize the learning of individual tasks, the DNN parameters are optimized to minimize the composite loss

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{MSE}} + \lambda \cdot \mathcal{L}_{\text{CrossEntropy}}, \qquad (25)$$

where the weight λ balances individual classification and regression tasks loss contributions since they might have considerably different values [228]. The reported results show that dual functionality can be implemented at a low additional computational cost.

Fig. 16 illustrates a generic dual-task learning framework to jointly localize the TX and classify its signal. The supervised learning requires labeled training data where each input frame s_{RX} has attached labels of waveform class and 2D coordinates $\{s_{\text{RX}}, \mathcal{H}_{\text{TX}}, x_{\text{TX}}, y_{\text{TX}}\}$. Input data s_{RX} in the original work [227] consists of raw IQ samples, however, reduced dimensionality representations could be an option to improve the algorithm complexity and inference time. The shared layers are shown as a black box since their configuration can vary significantly and their precise architecture is chosen empirically[8], the DNN outputs the estimated TX coordinates in 2D space $\hat{x}_{\text{TX}}, \hat{y}_{\text{TX}}$ (Output Layer 1) and the TX waveform class $\hat{\mathcal{H}}_{\text{TX}}$ (Output Layer 2), where four neurons correspond to the number of waveform classes in the dataset. The introduced architecture assumes the presence of a TX. Note that it can be further extended and trained to identify pure Noise input by simply adding another node to the Output Layer 2. Since this potential extension can be interpreted as spectrum sensing functionality, the NN becomes a triple-task system.

---

[8]In [227], the input data frame comprises the stacked vectors of 256 I and Q samples recorded by sensing units. The shared layers are represented by 4 LSTM and 2 fully connected layers.
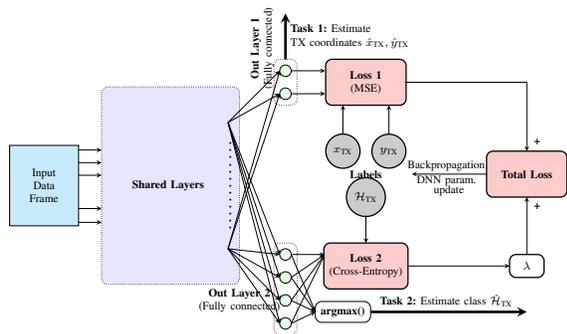
Fig. 16. Overview of the ML-based framework for dual-task learning for 4 waveform classes proposed in [227].
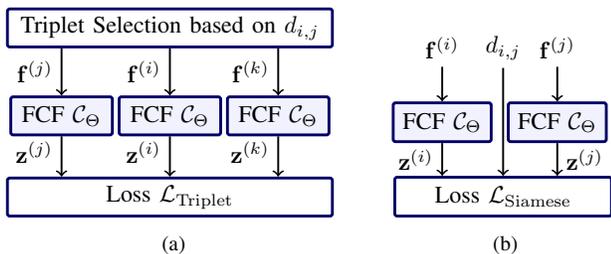

Fig. 17. Structure of Neural Networks commonly used for channel charting: (a) Triplet Neural Network and (b) Siamese Neural Network.

E. Channel Charting

Model-based TX positioning techniques operate on the premise of an assumed channel model, such as a LoS channel or a channel with limited multipath components. In contrast, data-based techniques explicitly exploit the unique characteristics of the wireless channel, often bypassing the need for a predefined model. Examples include fingerprinting, which leverages location-specific channel signatures, and channel charting [229], a novel method that maps high-dimensional CSI into a low-dimensional space while preserving geometric relationships.

Fingerprinting requires a reference positioning system to acquire labels for supervised training. Channel charting, on the other hand, is based on dimensionality reduction techniques to extract meaningful representations from high-dimensional CSI data. This process facilitates the creation of a "chart" where the relative positions of users in the environment are captured. Unlike fingerprinting, channel charting performs positioning via SSL without requiring absolute positioning information, where pseudo-labels are generated by the DL model and local positioning constraints are enforced to preserve the local geometry of the chart. The data-based techniques only rely on the assumption that there exists an injective mapping between the physical location of a TX and its associated channel, which is reasonable for sufficiently static environments.

The charting process typically involves training the so-called forward charting function (FCF) $\mathcal{C}_\Theta$ to identify and exploit intrinsic structures within the CSI data, making it more robust in complex environments. Often, channel charting relies on the notion of a *dissimilarity* between channels: Channels (each corresponding to a location in physical space) that are similar with respect to some dissimilarity metric (e.g., angle delay
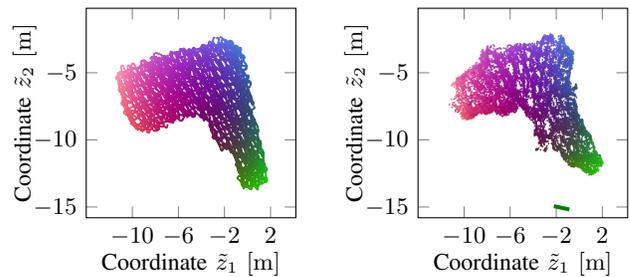
(a) Reference positions   (b) Learned channel chart and antenna array (green rectangle)

Fig. 18. Reference positions and learned channel chart in a top view chart. Measurement points are colorized and colors are preserved in the channel chart.

profile, timestamps), are said to have a low dissimilarity. If the FCF that performs the dimensionality reduction is realized as a NN, contrastive learning based on the dissimilarity can be applied. Two common NN structures for contrastive learning are the triplet NN shown in Fig. 17(a) and the Siamese NN shown in Fig. 17(b). In the case of the triplet NN, a triplet selection step presents three channel feature vectors to a weight-sharing NN: Two features vectors $\mathbf{f}^{(i)}$ and $\mathbf{f}^{(j)}$, which are considered to be similar to each other, and a third feature vector $\mathbf{f}^{(k)}$, which is considered to be dissimilar to $\mathbf{f}^{(i)}$ according to the dissimilarity metric. The triplet loss function is then designed to be minimal for channel charts where the low-dimensional latent space representations $\mathbf{z}^{(i)}$ and $\mathbf{z}^{(j)}$ are close to each other over all considered triplets (i, j, k), while also ensuring that $\mathbf{z}^{(k)}$ is far away from $\mathbf{z}^{(i)}$. For the Siamese NN, on the other hand, the loss function is designed to ensure that the Euclidean distance between any two latent space representations $\mathbf{z}^{(i)}$ and $\mathbf{z}^{(j)}$ in the Channel Chart matches the dissimilarity $d_{i,j}$ between those points as closely as possible. In general, the Triplet NN-based approach is more flexible and is compatible with various dissimilarity metrics. However, training the FCF in Siamese NN can lead to better performance, if a good dissimilarity metric is chosen [230].

An exemplary channel chart, taken from [231], is shown in Fig. 18 next to a chart of reference positions. Every point in the channel chart and reference position chart corresponds to one CSI measurement between a UE and a single antenna array with 2×4 antennas, operating at a carrier frequency of $1.272\,\mathrm{GHz}$ and a bandwidth of $50\,\mathrm{MHz}$. The CSI dataset is called *dichasus-cf0x* and taken from the DICHASUS dataset collection [232]. Even though large parts of the measurement are in a NLoS region for the antenna array or at least severely affected by multipath propagation, channel charting is able to reconstruct the global L-shaped topology of the environment so that the learned FCF can now be used for localization in both LoS and NLoS areas.

Channel charting can not only enable accurate user tracking and navigation, but also provides relative position estimates to the wireless network, which offers opportunities for improving network management. For example, pilots or spectral resources may be allocated based on user mobility patterns inferred from the channel chart.

F. Outlook and Lessons Learned

Passive radio sensing encompasses techniques such as spectrum sensing, signal classification, channel charting, and non-cooperative TX localization, all of which extract environmental or signal information without active transmissions. These methods rely on analyzing received signals to detect spectrum occupancy, identify waveform types, map radio propagation spaces, or estimate emitter positions. ML greatly enhances passive sensing by enabling systems to learn complex, nonlinear signal patterns and adapt to dynamic radio environments. Through data-driven feature extraction and inference, ML improves detection accuracy, classification robustness, and spatial awareness, making passive radio sensing more efficient, scalable, and resilient. Availability of the environment observations from geographically distributed nodes can greatly improve performance and even unlock some applications, and, therefore, is highly desirable. It can be assumed that in the future wireless networks, raw data recorded through opportunistic radio sensing will be used to realize multi-task learning, especially for the related tasks (see discussion in Sec. V). In this case, the bottleneck will be the need to collect a massive amount of data as well as GPU-accelerated platforms to process it in real-time.

V. EXPLORING AI-DRIVEN SYNERGIES BETWEEN DIFFERENT RADIO SENSING MODALITIES: OPPORTUNITIES AND IMPLEMENTATION ASPECTS

Motivated by the smart factory use case introduced in Section I-A, this section outlines potential benefits, architectural considerations, and key challenges for integrating radio sensing modalities discussed in Sections III and IV into a coherent framework for future wireless networks. We reinforce the significance of both the physical and electromagnetic environments' awareness, while also considering the major challenges and potential synergies between these domains. Fig. 19 illustrates several aspects of the AI-powered ISAC system presented in this section, and shows how system components and functionalities could be integrated in 6G.

A. Potential Benefits

It has been envisioned that in the future, there will be more integration between various sensing tasks [155]. By engaging different yet logically adjacent sensing modalities, the system can develop a holistic understanding of complex physical and electromagnetic environments and execute various tasks, thus fulfilling high expectations for future radio. Denoted as multi-modal multi-task learning (MM-MTL) [233], this paradigm is capable of simultaneously tackling tasks in a variety of domains and promising improved generalization and robustness. Below are a few possible examples of how MM-MTL can benefit wireless networks.

Remark. Note that there are other types of sensors which operate on fundamentally distinct principles, such as LIDARs and cameras. Just like radio sensing, the complementary strengths of those modalities show potential to enhance sensing accuracy, robustness, and generalization. In particular, these modalities are prominently used for autonomous driving, where multi-agent information fusion techniques employ diverse cooperation paradigms, sensor configurations, data sources, and application scenarios [234]. Another important technology, often considered as a separate sensing modality enabler, is the global positioning system (GPS). For instance, camera and GPS-assisted beam prediction was presented in [235] in the context of mmWave drone communication, where the data fusion helped in both improving beam prediction accuracy and considerably reducing the beam training overhead.

It is usually argued that radio sensing helps overcome limitations of visual data sensing; however, sometimes the opposite is true, and visual data can offset RF degradation under unfavorable conditions, such as multipath fading and interference. In our paper, however, we focus specifically on radio sensing.

1) Improved Security: Radio sensing is a cornerstone of network security. In a critical infrastructure setting (e.g., an airport surveillance system), radar could be used to detect the presence of unauthorized drones [13], while spectrum sensing identifies if those drones interfere by transmitting on the same frequencies the wireless network uses. The system could then adapt by switching the frequency or adjusting the transmit power to minimize the negative impact. Spectrum sensing alone can detect an interferer, but signal classification can provide additional insights about interference type and source. Furthermore, after detection, radar and non-cooperative TX localization techniques can complement each other to pinpoint the drone. Consequently, the wireless system might initialize counter-measures (e.g., removal) against the detected object. This synergetic approach can be applied to other scenarios, e.g., industrial monitoring systems detecting unauthorized sources.

2) Optimized Waveform Representation for Sensing: AI can not only help to perform the sensing task itself, but also identify the optimal waveform representation for specific sensing problems under certain conditions. For instance, it can help choose the minimum acceptable number of IQ samples for signal representation for raw data-driven AMC task, given the SINR conditions and the required accuracy. Furthermore, in some cases (e.g., low SINR), AI might decide to switch to statistical feature-based signal representation due to its potentially higher robustness, aiming to balance computational burden and traffic load according to accuracy requirements.

3) AI-Empowered Multi-Modal Radio Sensing as a Self-Sustaining Network Enabler: Early vision papers predicted that 6G will be a self-sustaining network (SSN) capable of continuously maintaining its key performance indicators (KPIs) under dynamic and complex environments [236]. Realization of the SSN concept can only rely on a high volume of environmental data, which can be provided via multi-modal sensing, while AI algorithms ensure its optimal utilization.

The network's awareness of both the physical environment and electromagnetic situation, potentially empowered by MM-MTL, enables SSNs as follows:

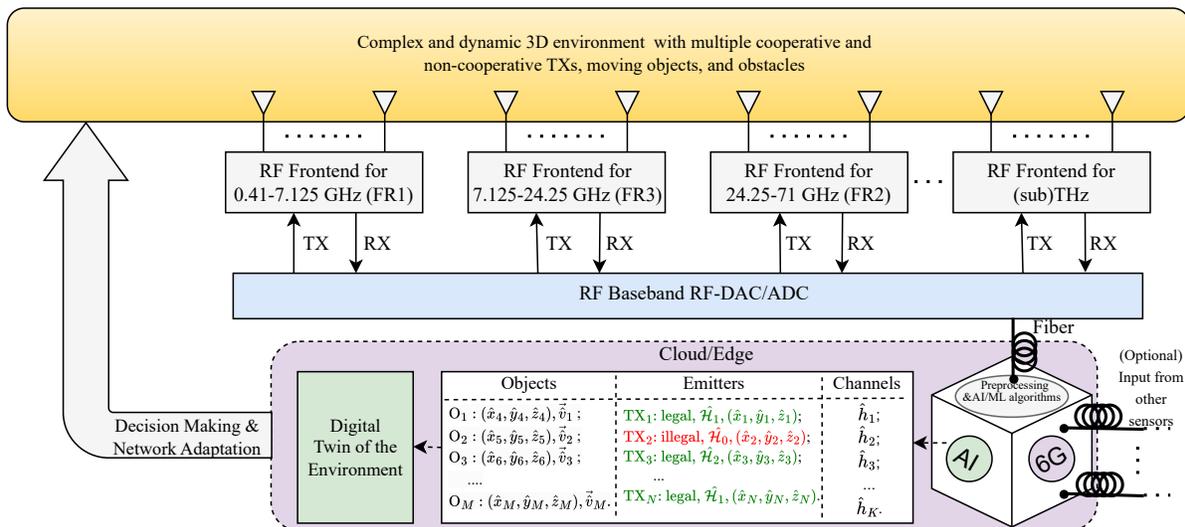- Radar can be used to detect blockers to adapt the beamforming.

Fig. 19. Illustration of the envisioned AI-enhanced multi-functional radio sensing integrated into the wireless network. We consider the dynamic environment with a complex propagation structure where multiple TX and passive objects are present. The sensing system is built on high-performance hardware, starting with the RF frontends, which are capable of scanning multiple channels in various FRs with an instantaneous bandwidth of a few GHz. RF baseband module is an FPGA which serves, among other things, as a DAC/ADC. In the RX mode, it outputs the complex baseband signals with high sampling frequency, which are subsequently fed via the fiber (or, alternatively, high-speed wireless backhaul links) to the cloud/edge computer, which performs the AI-based signal processing. Diverse information regarding the location and properties of the physical objects, TXs, and propagation structure, obtained through AI-based algorithms, is then used to build an ultimate DT of the environment. Such a DT facilitates the network's self-management functions, including interference management, beamforming, threat detection, etc. Incoming from the right side connections denote a system extension opportunity to implement distributed configurations, allowing data fusion algorithms.

- Spectrum sensing can help to identify idle spectrum, optimize channel assignment, and transmit power.
- Signal classification can leverage interference management, providing detailed information about TXs in the monitored area.
- Non-cooperative TX localization and channel charts can be used to improve the accuracy of positioning-based services as well as refine beamforming.

AI models can automatically optimize the network parameters, significantly increasing the degree of network autonomy. The concept of SSN aligns with DTs, where building a DT containing full knowledge of current EM and the physical environments can be seen as an intermediate step.

B. Implementation Aspects

Integrating the multi-band, multi-modal sensing functionality into future wireless networks will require tackling many practical issues. We assume a sensing-centric system, where communication is only used to exchange information regarding captured environmental data and those necessary for inference (e.g., parameters for federated learning (FL)-based model). The sensing can be realized as a dedicated multifunctional sensor module. Multifunctional sensors capable of spectrum sensing in the context of cognitive radar coexisting with communications networks were proposed in [163]. For the considered MM-MTL system, we take a step forward and imply a multifunctional sensor module as a platform integrating several sensing functionalities – for instance, radar, spectrum sensing, signal classification, non-cooperative TX localization, into an AI-powered dedicated sensing module. This module(s) could then be integrated into the radio access network, the facility surveillance system, etc., similarly to [11].

In the following, we outline hardware and real-time learning aspects of such a multi-task sensor module. In addition, we discuss architectural considerations and advocate for a distributed sensing approach to amplify the performance and reliability of this ambitious and complex hypothetic system.

1) Hardware: Bringing the multi-task ISAC to reality in a chaotic wireless environment will require advanced hardware components with a high degree of adaptivity. The hardware-related topics and challenges to be overcome for joint radar-communications are scrutinized in [13]. To also enable intelligent spectrum analysis functionality, high-performance RF frontends are essential since they provide algorithms with input raw data. This includes the need for wideband ADCs, RF amplifiers, and antenna arrays to capture a wide range of signal types. In particular, they should be capable of simultaneously tracking multiple frequencies and signal acquisition in multiple frequency bands, i.e., conversion of analog waveforms into digital form (baseband IQ samples) for storage and further processing by ML algorithms, as shown in Fig. 19. For example, an RF frontend based on direct sampling architecture from [237] is designed specifically for diverse sensing applications. It enables simultaneous sampling of all frequency bands below 6 GHz used in present-day mobile communications systems in Germany.[9]

In 6G networks, antennas will be reconfigurable, driving increased interest towards dynamic metasurface antennas (DMAs). DMAs can precisely and simultaneously control all key properties of electromagnetic waves, including amplitude, phase, frequency, polarization, etc. [238] This capability

---

[9]Although most standard radar frequency ranges are above 6 GHz, it is widely used, e.g., for long-range air traffic control and surveillance, long-range weather, and marine radar applications.

enables operation across wide frequency ranges, making them ideal for wideband spectrum sensing and diverse signal classification. Metasurfaces can act as selective filters, allowing specific signal types to pass through, simplifying signal classification based on spectral, temporal, or polarization characteristics. The tunable receiving patterns of DMAs are particularly valuable for passive sensing applications, while tunable radiation patterns are of interest for radar or channel sounding. A recently developed DMA proto-type, operating in the 60 GHz mmWave band, is presented in [239]. Combined in large numbers, such DMAs form a foundation for emerging holographic MIMO (HMIMO) receivers, spectrally agile structures capable of recording and manipulating electromagnetic fields with extreme flexibility and accuracy [240]. Enhanced with AI, HMIMO technology shows great potential for advancing 6G wireless applications, including futuristic intelligent spectrum monitoring and ISAC systems. A few alternative flexible-antenna systems include movable antennas [241], 6D movable antennas [242], and pinching antennas [243].

The bottleneck of many DL algorithms reviewed in Sections III and IV is that they often impose a heavy computational burden, especially those operating on high-dimensional raw data. It might cause problems like a slow training process and inference time, which then fail to meet real-time constraints. Leveraging GPUs and TPUs can help to meet real-time requirements of present and future wireless networks. For large capacity DNNs, GPUs and TPUs can accelerate the process by orders of magnitude. High-performance GPUs and TPUs are relatively expensive, leaving the question of practicality and affordability to a specific use case and budget constraints. Alternatives are the use of lightweight models or reduced dimensionality data representation. However, it usually results in performance loss. It is up to the particular application whether the accuracy is acceptable.

Besides high-performance components for wireless systems, an effort is underway to develop holistic testbeds for 6G and beyond research. For this purpose, some solutions aim not only to fulfill the communications requirements, but also specifically focus on enabling extra sensing features, such as distributed MIMO propagation measurements, radar target characterization, and ISAC [12], [244]. In particular, the broadband 6G MIMO testbed of the Open6GHub Project [12] offers a versatile hardware platform that supports multiple frequency bands for different use cases and scenarios, focusing on radar-based sensing and spectrum sensing. This testbed comprises high-performance hardware capable of wideband sampling and a backend server system that allows running scripts in MATLAB or Python in real time to test algorithms, including those AI-driven.

2) Architectural Aspects: Many algorithms described in Sections III and IV either explicitly use distributed ob-servations or could potentially benefit from them. Being integrated into a single node, sensing performance might suffer from blockage and other detrimental channel effects or the disability to monitor the whole area of interest due to lack of sensitivity, not to mention their potential breakdown. Moreover, some applications, such as blind TX localization,

are infeasible through a single observation. We therefore advocate for distributed and collaborative sensing since it helps to overcome these issues, increasing the accuracy, coverage, scalability, and robustness [49]. Concerning AI-powered networks, distributed sensing will naturally ensure more data for data-driven algorithm design.

The nodes of such a distributed sensing network might (*i*) comprise ubiquitous heterogeneous IoT devices with diverse sensing capabilities sharing the sensed data with an edge computer, (*ii*) be a network of high-performance dedicated modules at predefined locations (possibly, but not necessarily co-located with base stations or access points, or edge computers) explicitly designed to obtain spatial and spectrum awareness, or (*iii*) combine and balance contributions of both approaches, leveraging device diversity of (*i*) and high performance of individual nodes of (*ii*). With such a diversity of possible sensing data, the demand for diverse multimodal fusion models emerges. With the sensing processing at an edge computer possibly including computer vision tasks or the processing of many more different sensor observations, AI is a likely contender to solve these fusion tasks.

Among the key practical challenges of distributed AI-powered systems designed to simultaneously sense environ-ments and communicate data are latency, synchronization, and signaling overhead. The overall delay in ISAC originates from propagation delays and computational bottlenecks, e.g., inference time. Hence, a distributed architecture must also address how to balance the computational load and efficiently share information across the network to ensure real-time decision-making. Addressing multiple access issues is essen-tial to enable efficient communication and ensure reliable data transmission in dense, resource-limited networks [245]. Additionally, to reduce delay as well as energy consumption, data compression techniques can be used before the data is forwarded from one network node to another [246]. With this in mind, the role of AI is not limited to the inference of some sensing parameters. AI algorithms can optimize computational load across the network by analyzing the current load on network nodes and communication links.

Furthermore, a fundamental trade-off among latency, com-putational capacity, and privacy arises when one chooses between edge and cloud computing paradigms. Edge devices usually face constraints in computing power, memory, and energy [247]. Consequently, they use smaller models and compressed signal formats. In contrast, cloud-based inference enables very large models and centralized control, but it also introduces additional latency, dependency on reliable network connections, and a greater risk of RF data exposure. In this context, it is important to distinguish between *data privacy* and *physical privacy*. Data privacy refers to the protection of collected data (in the context of radio sensing, this might be IQ samples, spectral representations, or RDMs) and learned features during transmission or storage. This concern is most notable in cloud-based inference pipelines. Physical privacy, in contrast, stems from the inherent ability of radio sensing to extract sensitive information such as presence, location,

motion, activity patterns, etc.[10] For a comprehensive review of the current privacy-preserving techniques for protecting cyber-physical systems and their data from cyber-attacks, we refer reader to [248] and [249].

The concepts of FL and RL are highly important to ensure real-time adaptability in dynamic and unknown environments. RL is a natural choice for practical wireless applications since it is capable of learning while interacting with the environment in real-time. RL is widely used for radio resource allocation for real-time dynamic configurations in multi-device networks [250]. Early work that leverages RL to learn the whole MAC protocol in multi-agent networks for 6G networks was proposed in [251]; however, the drawback of this work is a relatively rigid centralized learning policy. Instead, [252] proposed a multi-agent deep RL (MADRL) framework adopting a fully distributed protocol design, reporting a superior performance in terms of throughput compared to current 5G NR-U for various traffic types, while offering adaptability and scalability over time. However, there are several obstacles that make the incorporation of online RL in practical wireless systems problematic. For instance, since RL is based on random exploration of the environment, the process can take too long, cause safety problems, and result in poor service provided to users [253]. In the context of sensing applications, addressing privacy concerns is particularly important. This issue is explicitly tackled via FL, where the local data recorded through sensing are not shared over the air. Furthermore, since FL assumes only sharing the model parameters, it reduces the stress over wireless links compared to raw data sharing. In wireless networks, choosing the devices that share their learned model parameters is a significant practical problem, as the principle 'the more, the better' is not always suitable. Over-the-air FL device scheduling has been widely accepted as an efficient sub-optimal solution, which reduces the performance degradation caused by imperfect model aggregation [254]. In brief, device scheduling aims to identify the set of devices that contribute to the aggregated model in the most constructive way, while turning the rest off. It is worth to mention that FL and RL can be used jointly to preserve data privacy and simultaneously adapt behavior in real-time [255].

Training a model on data from geographically distributed and potentially unsynchronized nodes can also be addressed by FL. In [256], various aspects of key IoT applications such as smart healthcare, smart transportation, UAVs, smart cities, and smart industry are discussed in the context of FL, including synchronization. Alternatively, time-series analysis with data from unsynchronized SUs can be utilized via post-processing techniques that estimate and correct the timestamps of data, or via algorithms that process aggregated data and do not require synchronization at all [227].

Finally, another aspect to consider is the NN architecture and the learning scheme for the sensing problem at hand. GNNs are particularly useful to capture the spatial interactions or interference between connected sensing and communication nodes. Additionally, GNNs possess unique features like

---

[10]Note that physical privacy risks remain even when all processing is performed on the device, since they stem from the sensing modality itself rather than data transmission.

permutation invariance and equivalence that ensure robust results under a different order of the sensing nodes. On the other hand, if a history of past measurements are available, transformers can exploit the temporal correlation between consecutive measurements to improve predicting the next state in transmission or parameter estimation. Regarding training of the NN, FL provides a privacy-preserving training approach for sensitive sensing tasks where the position of the objects or the user cannot be revealed through the transmission of raw data.

3) Adversarial Robustness: Adversarial robustness is a critical aspect of ML-based radio sensing, especially given the vulnerability of wireless environments to interference [257], spoofing [258], and general malicious signal injection [259]. Radio sensing systems heavily rely on often barely recognizable signal features that can easily be disturbed or mimicked by adversaries, which, depending on the task, might lead to misclassification, false detection, or degraded localization accuracy [260].

Adversarial attacks in ML-enhanced wireless systems typically work by adding carefully designed disturbances to the baseband IQ samples. Alternatively, such attacks might affect derived signal representations such as RDMs, micro-Doppler signatures, or CIRs. These disturbances can be introduced digitally, such as in stored datasets, or transmitted over the air with a low-power interfering waveform aligned with the target signal. Although these perturbations are usually kept small to maintain signal realism, they can still significantly change the features learned by neural networks. Common methods include gradient-based perturbations directly applied to IQ samples or feature-space attacks that target cyclostationary or time-frequency representations [261].

Defenses in radio sensing systems aim to make ML models more reliable while still preserving the physical properties of radio signals. A common method is adversarial training, where models are trained using both normal data and data that has been intentionally disturbed to simulate attacks. This can be done with a help of data augmentation such as random changes in phase or additive noise level [262]. Other approaches include detecting abnormal sensing results that do not match expected signal propagation or reflection behavior, e.g., targets suddenly appearing at impossible distances or speeds, reflections that jump randomly between range–Doppler bins. Together, these methods help reduce the model's sensitivity to small attacks while maintaining good performance in normal operating conditions.

A large proportion of the work on adversarial ML in wireless research is done in the context of AMC task. Some noticeable examples in recent literature include [263] and [264], where adversarial training was used to increase the classifier robustness. In the context of radar-communication systems, the need for enhanced security and robustness against adversarial ML attacks is investigated, e.g., in [265]. Specifically, the paper investigates adversarial attacks in ISAC systems for an automotive scenario, showing that attackers can inject false sensing information by manipulating RDM to mislead target detection. As a countermeasure, this work proposes a resilient adaptive ISAC framework that detects such attacks using a

GAN-based detector and dynamically adapts transmission modes to improve robustness and sensing accuracy.

Although adversarial ML and related training methods have been extensively studied in fields like computer vision and natural language processing, their effectiveness and adaptation for wireless sensing are still in the early stages [266]. We highlight adversarial robustness as one of the key research directions for the resilient deployment of ML-enhanced radio sensing systems in real-world scenarios and envision it to be a hot topic in the upcoming research.

4) Build Foundation Models: Building on the success of large language models (LLMs) across various domains, the telecommunications industry is now developing large telecom models (LTMs), domain-adapted foundation models specifically designed to address the unique demands of modern and future wireless networks [267], [268]. The concepts of foundation models and MM-MTL are closely related, and their design involves defining the use case, preparing large datasets, and choosing an appropriate architecture based on the data modality. Foundation models, with their ability to learn generalized representations from large, diverse datasets, can be trained to perform multiple radio sensing tasks, such as radar sensing, spectrum sensing, signal classification, and TX localization, within a unified framework. In the context of radio sensing, this could mean simultaneous learning from IQ samples, spectrograms, CIR, time-series metadata, spatial information, etc. Recent efforts in this direction include a self-supervised foundation model specifically tailored for wireless localization [269]. The designed model is suitable for ToA estimation, AoA estimation, single and multiple base station localization. Joint wireless technology recognition and localization based on input consisting of IQ samples and CIR timeseries is performed by a transformer-based foundation model in [270].

VI. CONCLUSION

Radio sensing and ML are two key intelligence enablers of future wireless networks, which are currently in the focus of the wireless research community and industry. Radio sensing aims to provide 6G wireless networks with a massive amount of raw data, which can be leveraged to extract the information—from type to location—about objects, TXs, and propagation environment. Meanwhile, ML-based algorithms will ensure unprecedented accuracy, autonomy, and efficiency of the networks. In this context, our paper provides a comprehensive overview of radio sensing as a key enabler for 6G, covering the necessary background in wireless and ML. It reviews ML-enhanced techniques for radar-based and passive electromagnetic sensing, discusses their potential integration into future networks, and highlights key challenges and design considerations.

While the vision of ML-powered radio sensing in 6G is promising, practical deployment faces several challenges. Many ML models remain too computationally intensive for wireless devices, calling for lightweight architectures optimized for real-time, low-power execution. On the other hand, to gain the full advantage of a large volume of sensed data and process it efficiently and in real time, AI hardware accelerators, such as specialized GPUs or dedicated AI chips, will be essential. Existing benchmark datasets used in research community are often outdated and do not reflect emerging technologies such as ISAC, thus limiting progress toward generalizable up-to-date solutions. At the same time, the rise of foundation models in the wireless community presents new opportunities. Their ability to generalize across tasks and modalities could significantly boost sensing capabilities, but their size, interpretability, and adaptability to wireless-specific constraints require further investigation. As AI becomes more central to wireless edge intelligence, careful integration with domain knowledge, attention to robustness, and a balance between distributed and centralized processing will be key to unlocking its full potential in future networks.

We hope that our work can serve as a starting point for novice researchers and engineers to better understand the state of the art, challenges, and possible research directions in the field of AI-enhanced radio sensing for future 6G wireless networks.

REFERENCES

[1] Y. LeCun, J. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 235–244, 2015.

[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

[3] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, 2012.

[4] J. Mitola, "Cognitive radio: An integrated agent architecture for software defined radio," *PhD dissertation*, 2000.

[5] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, 2005.

[6] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, 2017.

[7] S. Dörner, S. Cammerer, J. Hoydis, and S. ten Brink, "Deep learning based communication over the air," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 132–143, 2018.

[8] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality—what is next?: Five promising research directions for antenna arrays," *Digit. Signal Process.*, vol. 94, pp. 3–20, 2019.

[9] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, 2019.

[10] 3GPP TR 37.817, "Study on enhancement for data collection for NR and EN-DC," *V17.0.0*, 2022.

[11] V. Shatov, M. Lübke, Y. Su, and N. Franchi, "On integrated cooperative radio sensing for spatial electromagnetic analysis in 6G," in *Proc. IEEE Future Networks World Forum (FNWF)*, 2023, pp. 1–8.

[12] B. Nuss *et al.*, "Flexible and scalable broadband massive MIMO testbed for joint communication and sensing applications," in *Proc. European Wireless*, 2024, pp. 1–6.

[13] V. Shatov *et al.*, "Joint radar and communications: Architectures, use cases, aspects of radio access, signal processing, and hardware," *IEEE Access*, vol. 12, pp. 47 888–47 914, 2024.

[14] M. Hoffmann *et al.*, "A secure and resilient 6G architecture vision of the German flagship project 6G-ANNA," *IEEE Access*, vol. 11, pp. 102 643–102 660, 2023.

[15] A. Schösser, F. Burmeister, P. Schulz, M. D. Khursheed, S. Ma, and G. Fettweis, "Advancing spectrum anomaly detection through digital twins," *IEEE Commun. Mag.*, vol. 63, no. 4, pp. 40–46, 2025.

[16] A. Bazzi, R. Bomfin, M. Mezzavilla, S. Rangan, T. Rappaport, and M. Chafii, "Upper mid-band spectrum for 6G: Vision, opportunity and challenges," *IEEE Commun. Mag.*, vol. 64, no. 1, pp. 206–212, 2026.

[17] X. Lin, L. Kundu, C. Dick, and S. Velayutham, "Embracing AI in 5G-advanced toward 6G: A joint 3GPP and O-RAN perspective," *IEEE Commun. Stand. Mag.*, vol. 7, no. 4, pp. 76–83, 2023.

[18] X. Lin, "Artificial intelligence in 3GPP 5G-Advanced: A survey," *IEEE ComSoc Technology News*, 2023.

[19] X. Lin *et al.*, "AI-native 6G: Empowering intelligent RAN with accelerated compute," *IEEE Wirel. Commun.*, vol. 32, no. 6, pp. 11–14, 2025.

[20] 3GPP, "Feasibility study on integrated sensing and communication," *Tech. Rep. 22.837 (Version 19.3.0)*, 2024.

[21] International Telecommunication Union - Radiocommunication Sector (ITU-R), "Framework and overall objectives of the future development of IMT for 2030 and beyond," International Telecommunication Union, Recommendation ITU-R M.2160-0, 2023.

[22] N. Wu *et al.*, "AI-enhanced integrated sensing and communications: Advancements, challenges, and prospects," *IEEE Commun. Mag.*, vol. 62, no. 9, pp. 144–150, 2024.

[23] G. Cerro, G. Miele, R. Hislop, O. Holland, and A. Mody, "Ieee 802.15.22.3 spectrum characterization and occupancy sensing application testbed," *IEEE Trans. Instrum. Meas.*, vol. 23, no. 9, pp. 58–64, 2020.

[24] IEEE. (2019) 802.22, spectrum characterization and occupancy sensing work group. [Online]. Available: https://www.ieee802.org/22/HomePageText.htm

[25] Z. Wei *et al.*, "Integrated sensing and communication driven digital twin for intelligent machine network," *IEEE Internet Things Mag.*, vol. 7, no. 4, pp. 60–67, 2024.

[26] Nokia. (2024) 6G explained. [Online]. Available: https://www.nokia.com/about-us/newsroom/articles/6g-explained/

[27] Ericsson. (2024) White paper: Co-creating a cyber-physical world. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/white-papers/co-creating-a-cyber-physical-world

[28] T. Wen, M. Jianglei, Z. Peiying, and C. Yan. (2024) Co-creating a cyber-physical world. [Online]. Available: https://www.huawei.com/en/huaweitech/publication/202401/ai-bridge-to-6g#

[29] E. Obiodu, K. Chowdhury, L. Kundu, and X. Lin. (2024) Boosting AI-driven innovation in 6G with the AI-RAN Alliance, 3GPP, and O-RAN. [Online]. Available: https://developer.nvidia.com/blog/boosting-ai-driven-innovation-in-6g-with-the-ai-ran-alliance-3gpp\-and-o-ran/

[30] S. Cammerer *et al.*, "Sionna research kit: A GPU-accelerated research platform for AI-RAN," in *Proc. IEEE Int. Conf. on Machine Learning for Commun. and Netw. (ICMLCN)*, 2025.

[31] Nokia. (2025) Nokia partners with Nvidia. [Online]. Available: https://www.nokia.com/newsroom/nokia-partners-with-nvidia/

[32] A. He *et al.*, "A survey of artificial intelligence for cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 59, no. 4, pp. 1578–1592, 2010.

[33] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine-learning techniques in cognitive radios," *IEEE Commun. Surv. Tutor.*, vol. 15, no. 3, pp. 1136–1159, 2013.

[34] C. De Lima *et al.*, "Convergent communication, sensing and localization in 6G systems: An overview of technologies, opportunities and challenges," *IEEE Access*, vol. 9, pp. 26 902–26 925, 2021.

[35] F. Liu *et al.*, "Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1728–1767, 2022.

[36] U. Demirhan and A. Alkhateeb, "Integrated sensing and communication for 6G: Ten key machine learning roles," *IEEE Commun. Mag.*, vol. 61, no. 5, pp. 113–119, May 2023.

[37] M. Ade Krisna Respati and B. M. Lee, "A survey on machine learning enhanced integrated sensing and communication systems: Architectures, algorithms, and applications," *IEEE Access*, vol. 12, pp. 170 946–170 964, 2024.

[38] N. A. Khalek, D. H. Tashman, and W. Hamouda, "Advances in machine learning-driven cognitive radio for wireless networks: A survey," *IEEE Commun. Surv. Tutor.*, vol. 26, no. 2, pp. 1201–1237, 2024.

[39] P. Lohan, B. Kantarci, M. Amine Ferrag, N. Tihanyi, and Y. Shi, "From 5G to 6G networks: A survey on AI-based jamming and interference detection and mitigation," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 3920–3974, 2024.

[40] N. González-Prelcic *et al.*, "The integrated sensing and communication revolution for 6G: Vision, techniques, and applications," *Proc. IEEE*, vol. 112, no. 7, pp. 676–723, 2024.

[41] S. Lu *et al.*, "Integrated sensing and communications: Recent advances and ten open challenges," *IEEE Internet Things J.*, vol. 11, no. 11, pp. 19 094–19 120, 2024.

[42] S. N. Syed *et al.*, "Deep neural networks for spectrum sensing: A review," *IEEE Access*, vol. 11, pp. 89 591–89 615, 2023.

[43] F. R. V. Guimarães, J. M. B. da Silva Jr., C. Casimiro Cavalcante, G. Fodor, M. Bengtsson, and C. Fischione, "Machine learning for spectrum sharing: A survey," *Foundations and Trends in Networking*, vol. 14, no. 1–2, p. 1–159, 2024.

[44] D. Wen, Y. Zhou, X. Li, Y. Shi, K. Huang, and K. B. Letaief, "A survey on integrated sensing, communication, and computation," *IEEE Commun. Surv. Tutor.*, vol. 27, no. 5, pp. 3058–3098, 2025.

[45] X. Cao *et al.*, "AI-empowered multiple access for 6G: A survey of spectrum sensing, protocol designs, and optimizations," *Proc. IEEE*, vol. 112, no. 9, pp. 1264–1302, 2024.

[46] E. C. Strinati *et al.*, "Toward distributed and intelligent integrated sensing and communications for 6G networks," *IEEE Wirel. Commun.*, vol. 32, no. 1, pp. 60–67, 2025.

[47] X. Luo, Q. Lin, R. Zhang, H.-H. Chen, X. Wang, and M. Huang, "ISAC—a survey on its layered architecture, technologies, standardizations, prototypes, and testbeds," *IEEE Commun. Surv. Tutor.*, vol. 28, pp. 485–526, 2026.

[48] N. C. Luong *et al.*, "Advanced learning algorithms for integrated sensing and communication (ISAC) systems in 6G and beyond: A comprehensive survey," *IEEE Commun. Surv. Tutor.*, vol. 28, pp. 2572–2611, 2026.

[49] E. C. Strinati *et al.*, "Distributed intelligent integrated sensing and communications: The 6G-DISAC approach," in *Proc. Joint European Conf. on Netw. and Commun. & 6G Summit (EuCNC/6G Summit)*, 2024, pp. 392–397.

[50] G. L. Stuber, *Principles of Mobile Communication*, 4th ed. Springer Nature, 2017.

[51] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[52] T. Rappaport, *Wireless communications: Principles and practice*, 2nd ed., ser. Prentice Hall communications engineering and emerging technologies series. Prentice Hall, 2002.

[53] H. L. van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. Wiley, 2002.

[54] A. S. Rahmathullah, Á. F. García-Fernández, and L. Svensson, "Generalized optimal sub-pattern assignment metric," in *2017 20th International Conference on Information Fusion (Fusion)*. IEEE, 2017, pp. 1–8.

[55] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., 1993.

[56] S. Padakandla, "A survey of reinforcement learning algorithms for dynamically varying environments," *ACM Comput. Surv.*, vol. 54, no. 6, Jul. 2021. [Online]. Available: https://doi.org/10.1145/3459991

[57] Y. LeCun and Y. Bengio, *Convolutional networks for images, speech, and time series*. Cambridge, MA, USA: MIT Press, 1998, p. 255–258.

[58] A. Vaswani *et al.*, "Attention is all you need," in *Proc. of the 31st Adv. Neural Inf. Process. (NeurIPS)*. Curran Associates Inc., 2017, p. 6000–6010.

[59] G. Corso, H. Stark, S. Jegelka, T. Jaakkola, and R. Barzilay, "Graph neural networks," *Nature Reviews Methods Primers*, vol. 4, no. 1, p. 17, 2024.

[60] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, 2021.

[61] ETSI 3rd Generation Partnership Project (3GPP), "Study on channel model for frequencies from 0.5 to 100 GHz (3GPP TR 38.901 version 16.1.0 Release 16)," ETSI, Tech. Rep., 2020.

[62] F. Burkhardt, S. Jaeckel, E. Eberlein, and R. Prieto-Cerdeira, "QuaDRiGa: a MIMO channel model for land mobile satellite," in *Proc. 8th European Conf. on Antennas and Propag. (EuCAP 2014)*. IEEE, 2014, pp. 1274–1278.

[63] J. Hoydis *et al.*, "Sionna RT: Differentiable ray tracing for radio propagation modeling," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, 2023, pp. 317–321.

[64] Remcom, "Wireless InSite," http://www.remcom.com/wireless-insite.

[65] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," in *Proc. Information Theory and Applications Workshop (ITA)*, San Diego, CA, Feb 2019, pp. 1–8.

[66] National Institute of Standards and Technology (NIST), "NIST NextG Channel Model Alliance," 2025, accessed: 2025-02-13. [Online]. Available: https://nextg.nist.gov/

[67] O. Kanhere, H. Poddar, and T. S. Rappaport, "Calibration of NYURay for ray tracing using 28, 73, and 142 GHz channel measurements

[67] conducted in indoor, outdoor, and factory scenarios," *IEEE Trans. Antennas Propag.*, vol. 73, no. 1, pp. 405–420, 2025.

[68] N. Benvenuto and F. Piazza, "On the complex backpropagation algorithm," *IEEE Trans. Signal Process.*, vol. 40, no. 4, pp. 967–969, 1992.

[69] N. Shlezinger, J. Whang, Y. C. Eldar, and A. G. Dimakis, "Model-based deep learning," *Proc. IEEE*, vol. 111, no. 5, pp. 465–499, 2023.

[70] W. Jiang, D. Ma, Z. Wei, Z. Feng, P. Zhang, and J. Peng, "ISAC-NET: Model-driven deep learning for integrated passive sensing and communication," *IEEE Trans. Commun.*, vol. 72, no. 8, pp. 4692–4707, 2024.

[71] J. K. Eshraghian *et al.*, "Training spiking neural networks using lessons from deep learning," *IEEE Proc.*, vol. 111, no. 9, pp. 1016–1054, Sep. 2023.

[72] J. Chen, N. Skatchkovsky, and O. Simeone, "Neuromorphic integrated sensing and communications," *IEEE Wirel. Commun. Lett.*, vol. 12, no. 3, pp. 476–480, 2023.

[73] R. S. Thoma *et al.*, "Cooperative passive coherent location: A promising 5g service to support road safety," *IEEE Commun. Mag.*, vol. 57, no. 9, pp. 86–92, 2019.

[74] Y. Xiong, F. Liu, Y. Cui, W. Yuan, T. X. Han, and G. Caire, "On the fundamental tradeoff of integrated sensing and communications under gaussian channels," *IEEE Trans. Inf. Theory*, vol. 69, no. 9, pp. 5723–5751, 2023.

[75] L. Yang, Y. Wei, Z. Feng, Q. Zhang, and Z. Han, "Deep reinforcement learning-based resource allocation for integrated sensing, communication, and computation in vehicular network," *IEEE Trans. Wirel. Commun.*, vol. 23, no. 12, pp. 18 608–18 622, Dec. 2024.

[76] C. Liu, M. Xia, J. Zhao, H. Li, and Y. Gong, "Optimal resource allocation for integrated sensing and communications in internet of vehicles: A deep reinforcement learning approach," *IEEE Trans. Veh. Technol.*, vol. 74, no. 2, pp. 3028–3038, Feb. 2025.

[77] Y. Qin, Z. Zhang, X. Li, W. Huangfu, and H. Zhang, "Deep reinforcement learning based resource allocation and trajectory planning in integrated sensing and communications uav network," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 8158–8169, 2023.

[78] X. Ye, Y. Mao, X. Yu, S. Sun, L. Fu, and J. Xu, "Integrated sensing and communications for low-altitude economy: A deep reinforcement learning approach," *IEEE Trans. Wireless Commun.*, 2025, early access.

[79] C. Muth, B. Geiger, D. G. Gaviria, and L. Schmalen, "Loss design for single-carrier joint communication and neural network-based sensing," in *Proc. 27th Int. Workshop on Smart Antennas (WSA)*, 2024, pp. 131–137.

[80] J. M. Mateos-Ramos *et al.*, "End-to-end learning for integrated sensing and communication," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2022.

[81] C. Muth and L. Schmalen, "Autoencoder-based joint communication and sensing of multiple targets," in *Proc. 26th Int. ITG Workshop on Smart Antennas and 13th Conf. on Systems, Communications, and Coding (WSA & SCC)*, 2023, pp. 1–6.

[82] G. Fontanesi *et al.*, "A deep-NN beamforming approach for dual function radar-communication THz UAV," *IEEE Trans. Veh. Technol.*, vol. 74, no. 1, pp. 746–760, 2025.

[83] N. T. Nguyen, L. V. Nguyen, N. Shlezinger, Y. C. Eldar, A. L. Swindlehurst, and M. Juntti, "Joint communications and sensing hybrid beamforming design via deep unfolding," *IEEE J. Sel. Top. Signal Process.*, vol. 18, no. 5, pp. 901–916, 2024.

[84] Y. Zhang, S. Li, D. Li, J. Zhu, and Q. Guan, "Transformer-based predictive beamforming for integrated sensing and communication in vehicular networks," *IEEE Internet Things J.*, vol. 11, no. 11, pp. 20 690–20 705, 2024.

[85] J. Lee, Y. Cheng, D. Niyato, Y. L. Guan, and D. González G., "Intelligent resource allocation in joint radar-communication with graph neural networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 10, pp. 11 120–11 135, 2022.

[86] X. Li, M. Chen, Y. Liu, Z. Zhang, D. Liu, and S. Mao, "Graph neural networks for joint communication and sensing optimization in vehicular networks," *IEEE J. Sel. Areas Commun*, vol. 41, no. 12, pp. 3893–3907, 2023.

[87] Y. Zhao, Y. Zhou, Z. Wang, Y. Shi, N. Cheng, and H. Zhou, "Learning to beamform for integrated sensing and communication: A graph neural network with implicit projection approach," *IEEE Trans. Wireless Commun.*, vol. 24, no. 7, pp. 5931–5945, 2025.

[88] X. Yang, R. Zhang, D. Zhai, F. Liu, R. Du, and T. X. Han, "Constellation design for integrated sensing and communication with random waveforms," *IEEE Trans. Wirel. Commun.*, vol. 23, no. 11, pp. 17 415–17 428, 2024.

[89] B. Geiger, F. Liu, S. Lu, A. Rode, and L. Schmalen, "Joint optimization of geometric and probabilistic constellation shaping for OFDM-ISAC systems," in *Proc. IEEE 5th Int. Symp. on Joint Commun. & Sensing (JC&S)*, 2025, pp. 1–6.

[90] W. Jiang, A. M. Haimovich, and O. Simeone, "End-to-end learning of waveform generation and detection for radar systems," in *Proc. 53rd Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, 2019, pp. 1672–1676.

[91] B. Major *et al.*, "Vehicle detection with automotive radar using deep learning on range-azimuth-doppler tensors," in *Proc. IEEE/CVF Int. Conf. on Computer Vision Workshop (ICCVW)*, 2019, pp. 924–932.

[92] H. Dong and O. B. Akan, "Debrisense: Terahertz-based integrated sensing and communications (isac) for debris detection and classification in the internet of space (ios)," 2024.

[93] P. Tosi, S. Schieler, M. Henninger, S. Semper, and S. Mandelli, "Benchmarking CFAR and CNN-based peak detection algorithms in ISAC under hardware impairments," in *Proc. 28th Int. Workshop on Smart Antennas (WSA)*, 2025.

[94] S. Schieler, S. Semper, and R. Thomä, "Wireless propagation parameter estimation with convolutional neural networks," *Int. J. Microw. Wirel. Technol.*, vol. 17, no. 5, p. 892–899, 2025.

[95] J. Beuster *et al.*, "Measurement testbed for radar and emitter localization of uav at 3.75 ghz," in *2023 17th European Conf. on Antennas and Propag. (EuCAP)*, 2023, pp. 1–5.

[96] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2016. [Online]. Available: https://arxiv.org/abs/1506.02640

[97] S. Schieler, S. Semper, and R. Thomä, "Wireless propagation parameter estimation with convolutional neural networks," *Int. J. Microw. Wirel. Technol.*, vol. 17, no. 5, p. 892–899, 2025.

[98] A. Richter, *Estimation of radio channel parameters: models and algorithms.* Ilmenau: ISLE, 2005.

[99] S. Semper, J. Naviliat, J. Gedschold, M. Döbereiner, S. Schieler, and R. S. Thomä, "Distributed computing and model-based estimation for integrated communications and sensing: A roadmap," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 6279–6290, 2024.

[100] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Trans. Signal Process.*, vol. 39, no. 11, pp. 2436–2449, 1991.

[101] P. Stoica and K. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoustics, Speech, and Signal Process.*, vol. 38, no. 7, pp. 1132–1143, 1990.

[102] K. Sharman and G. McClurkin, "Genetic algorithms for maximum likelihood parameter estimation," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Process.*, 1989, pp. 2716–2719 vol.4.

[103] S. Zheng *et al.*, "Deep learning-based DOA estimation," *IEEE Trans. Cogn. Commun. Netw.*, vol. 10, no. 3, pp. 819–835, 2024.

[104] Y. Ma, Y. Zeng, and S. Sun, "A deep learning based super resolution DoA estimator with single snapshot MIMO radar data," *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 4142–4155, 2022.

[105] P. Chen, Z. Chen, L. Liu, Y. Chen, and X. Wang, "Sdoa-net: An efficient deep-learning-based DOA estimation network for imperfect array," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–12, 2024.

[106] W. Zhu, M. Zhang, P. Li, and C. Wu, "Two-dimensional DOA estimation via deep ensemble learning," *IEEE Access*, vol. 8, pp. 124 544–124 552, 2020.

[107] D. Chen, S. Shi, X. Gu, and B. Shim, "Robust DoA estimation using denoising autoencoder and deep neural networks," *IEEE Access*, vol. 10, pp. 52 551–52 564, 2022.

[108] J. Yu and Y. Wang, "Deep learning-based multipath DoAs estimation method for mmwave massive MIMO systems in low SNR," *IEEE Trans. Veh. Technol.*, vol. 72, no. 6, pp. 7480–7490, 2023.

[109] D. H. Shmuel, J. P. Merkofer, G. Revach, R. J. G. v. Sloun, and N. Shlezinger, "Subspacenet: Deep learning-aided subspace methods for DoA estimation," *IEEE Trans. Veh. Technol.*, pp. 1–15, 2024.

[110] J. P. Merkofer, G. Revach, N. Shlezinger, T. Routtenberg, and R. J. G. van Sloun, "DA-MUSIC: Data-driven DoA estimation via deep augmented MUSIC algorithm," *IEEE Trans. Veh. Technol.*, vol. 73, no. 2, pp. 2771–2785, 2024.

[111] B. Chatelier *et al.*, "Physically parameterized differentiable music for doa estimation with uncalibrated arrays," *IEEE Int. Conf. on Commun.*, 2025.

[112] A. Barthelme and W. Utschick, "A machine learning approach to DoA estimation and model order selection for antenna arrays with subarray sampling," *IEEE Trans. Signal Process.*, vol. 69, pp. 3075–3087, 2021.

[113] S. Schieler, S. Semper, R. Faramarzahangari, M. Döbereiner, C. Schneider, and R. Thomä, "Grid-free harmonic retrieval and model order selection using convolutional neural networks," in *Proc. 18th European Conf. on Antennas and Propag. (EuCAP)*, 2024, pp. 1–5.

[114] S. Schieler, S. Semper, C. Schneider, and R. Thomä, "Measurement-based evaluation of cnn-based detection and estimation for ISAC systems," in *Proc. IEEE Int. Radar Conf.*, 2025, pp. 1–5.

[115] A. Zyweck and R. Bogner, "Radar target classification of commercial aircraft," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 2, pp. 598–606, 1996.

[116] P. Sorowka and H. Rohling, "Pedestrian classification with 24 GHz chirp sequence radar," in *Proc. 16th Int. Radar Symp. (IRS)*, 2015, pp. 167–173.

[117] S. Gupta, P. K. Rai, A. Kumar, P. K. Yalavarthy, and L. R. Cenkeramaddi, "Target classification by mmwave FMCW radars using machine learning on range-angle images," *IEEE Sens. J.*, vol. 21, no. 18, pp. 19 993–20 001, 2021.

[118] M. Ulrich, C. Gläser, and F. Timm, "Deepreflecs: Deep learning for automotive object classification with radar reflections," in *Proc. IEEE Radar Conf. (RadarConf21)*, 2021, pp. 1–6.

[119] K. Patel, K. Rambach, T. Visentin, D. Rusev, M. Pfeiffer, and B. Yang, "Deep learning-based object classification on automotive radar spectra," in *Proc. IEEE Radar Conf. (RadarConf)*, 2019, pp. 1–6.

[120] J. M. Mateos-Ramos, C. Häger, M. F. Keskin, L. L. Magoarou, and H. Wymeersch, "Model-based end-to-end learning for multi-target integrated sensing and communication under hardware impairments," *IEEE Trans. Wirel. Commun.*, pp. 2574–2589, 2025.

[121] J. M. Mateos-Ramos, B. Chatelier, C. Häger, L. Le Magoarou, and H. Wymeersch, "Semi-supervised end-to-end learning for integrated sensing and communications," in *Proc. IEEE Int. Conf. on Machine Learning for Commun. and Netw. (ICMLCN)*, 2024, pp. 132–138.

[122] K. Hu, X. Hu, L. Qi, G. Lu, Y. Zhong, and Y. Han, "Radnet: A radar detection network for target detection using 3D range-angle-doppler tensor," in *Proc. IEEE 19th Int. Conf. on Automation Science and Engineering (CASE)*, 2023, pp. 1–6.

[123] Z. Benyahia, M. Hefnawi, M. Aboulfatah, E. Abdelmounim, and T. Gadi, "Squeezenet-based range, angle, and doppler estimation for automotive MIMO radar systems," in *Proc. Int. Conf. on Intelligent Systems and Computer Vision (ISCV)*, 2022, pp. 1–5.

[124] S. Schieler, S. Semper, R. Faramarzahangari, C. Schneider, and R. S. Thomae, "4d joint harmonic retrieval and model order estimation with convolutional neural networks," in *Proc. 5th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2023)*. IFSA Publishing, S. L., 2023, pp. 204–208. [Online]. Available: \url{https://sensorsportal.com/ASPAI_2023/ASPAI_2023_Conference_Proceedings.pdf}

[125] Á. F. García-Fernández and L. Svensson, "Trajectory PHD and CPHD filters," *IEEE Trans. Signal Process.*, vol. 67, no. 22, pp. 5702–5714, 2019.

[126] B.-N. Vo and B.-T. Vo, "A multi-scan labeled random finite set model for multi-object state estimation," *IEEE Trans. Signal Process.*, vol. 67, no. 19, pp. 4948–4963, 2019.

[127] Y. Xia, K. Granström, L. Svensson, Á. F. García-Fernández, and J. L. Williams, "Multi-scan implementation of the trajectory poisson multi-bernoulli mixture filter," *Journal of Advances in Information Fusion*, vol. 14, no. 2, pp. 213–235, 2019.

[128] R. Mahler, *Statistical multisource-multitarget information fusion*. Artech, 2007.

[129] C. Liu *et al.*, "Learning-based predictive beamforming for integrated sensing and communication in vehicular networks," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 8, pp. 2317–2334, 2022.

[130] Z. Wang and V. W. Wong, "Deep learning for ISAC-enabled end-to-end predictive beamforming in vehicular networks," in *Proc. IEEE Int. Conf. on Commun.*, 2023, pp. 5713–5718.

[131] U. Demirhan and A. Alkhateeb, "Radar aided 6G beam prediction: Deep learning algorithms and real-world demonstration," in *Proc. IEEE Wirel. Commun. and Netw. Conf. (WCNC)*, 2022, pp. 2655–2660.

[132] J. Pinto, G. Hess, W. Ljungbergh, Y. Xia, H. Wymeersch, and L. Svensson, "Deep learning for model-based multiobject tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 6, pp. 7363–7379, 2023.

[133] J. Pinto, G. Hess, Y. Xia, H. Wymeersch, and L. Svensson, "Transformer-based multi-object smoothing with decoupled data association and smoothing," *arXiv preprint arXiv:2312.17261*, 2023.

[134] M. Gruber, *An approach to target tracking*. MIT Lincoln Laboratory, 1967.

[135] S. J. Julier and J. K. Uhlmann, "New extension of the kalman filter to nonlinear systems," in *Signal processing, sensor fusion, and target recognition VI*, vol. 3068. Spie, 1997, pp. 182–193.

[136] N. J. Gordon, D. J. Salmond, and A. F. Smith, "Novel approach to nonlinear/non-gaussian bayesian state estimation," in *IEE proceedings F (radar and signal processing)*, vol. 140, no. 2. IET, 1993, pp. 107–113.

[137] G. Revach, N. Shlezinger, X. Ni, A. L. Escoriza, R. J. G. van Sloun, and Y. C. Eldar, "KalmanNet: Neural network aided kalman filtering for partially known dynamics," *IEEE Trans. Signal Process.*, vol. 70, pp. 1532–1547, 2022.

[138] X. Ni, G. Revach, and N. Shlezinger, "Adaptive kalmannet: Data-driven Kalman filter with fast adaptation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2024, pp. 5970–5974.

[139] Y. Dahan, G. Revach, J. Dunik, and N. Shlezinger, "Uncertainty quantification in deep learning based kalman filters," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2024, pp. 13 121–13 125.

[140] I. Nuri and N. Shlezinger, "Learning flock: Enhancing sets of particles for multi˜ sub-state particle filtering with neural augmentation," *arXiv preprint arXiv:2408.11348*, 2024.

[141] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 3, pp. 311–324, 2007.

[142] N. Lu, Y. Wu, L. Feng, and J. Song, "Deep learning for fall detection: Three-dimensional CNN combined with LSTM on video kinematic data," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 1, pp. 314–323, 2019.

[143] S. C. Mukhopadhyay, "Wearable sensors for human activity monitoring: A review," *IEEE Sens. J.*, vol. 15, no. 3, pp. 1321–1330, 2015.

[144] W. Qi, R. Zhang, J. Zhou, H. Zhang, Y. Xie, and X. Jing, "A resource-efficient cross-domain sensing method for device-free gesture recognition with federated transfer learning," *IEEE Trans. Green Commun. Netw*, vol. 7, no. 1, pp. 393–400, 2023.

[145] H. Du, T. Jin, Y. Song, Y. Dai, and M. Li, "A three-dimensional deep learning framework for human behavior analysis using range-doppler time points," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 611–615, 2020.

[146] N. N. Bhat, R. Berkvens, and J. Famaey, "Gesture recognition with mmwave Wi-Fi access points: Lessons learned," in *Proc. IEEE 24th Int. Symp. on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2023, pp. 127–136.

[147] Y. Zhao, A. Yarovoy, and F. Fioranelli, "Angle-insensitive human motion and posture recognition based on 4D imaging radar and deep learning classifiers," *IEEE Sens. J.*, vol. 22, no. 12, pp. 12 173–12 182, 2022.

[148] J. Hoydis *et al.*, "Learning radio environments by differentiable ray tracing," *IEEE Trans. on Machine Learning in Commun. and Netw.*, vol. 2, pp. 1527–1539, 2024.

[149] V. Bouget, D. Béréziat, J. Brajard, A. Charantonis, and A. Filoche, "Fusion of rain radar images and wind forecasts in a deep learning model applied to rain nowcasting," *Remote Sensing*, vol. 13, no. 2, p. 246, 2021.

[150] A. Ferdowsi and D. Whitefield, "Deep learning for rain fade prediction in satellite communications," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, 2021, pp. 1–6.

[151] S. Helal, H. Sarieddeen, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Signal processing and machine learning techniques for terahertz sensing: An overview," *IEEE Signal Process. Mag.*, vol. 39, no. 5, pp. 42–62, sep 2022.

[152] R. Thomä and T. Dallmann, "Distributed ISAC systems – multisensor radio access and coordination," in *Proc. 20th European Radar Conf. (EuRAD)*, 2023, pp. 351–354.

[153] J. Guo, J. Yang, T. Du, C.-K. Wen, and S. Jin, "Semantic-aware monostatic ISAC system with dynamic multimodal fusion," in *2025 IEEE/CIC International Conference on Communications in China (ICCC)*, 2025, pp. 1–6.

[154] Z. Zhuang, D. Wen, Y. Shi, G. Zhu, S. Wu, and D. Niyato, "Integrated sensing-communication-computation for over-the-air edge AI inference," *IEEE Trans. Wirel. Commun.*, vol. 23, no. 4, pp. 3205–3220, 2024.

[155] X. Cheng *et al.*, "Intelligent multi-modal sensing-communication integration: Synesthesia of machines," *IEEE Commun. Surv. Tutor.*, vol. 26, no. 1, pp. 258–301, 2024.

[156] T. Meng, X. Jing, Z. Yan, and W. Pedrycz, "A survey on machine learning for data fusion," *Information Fusion*, vol. 57, pp. 115–129,

2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253519303902

[157] A. Alkhateeb *et al.*, "DeepSense 6G: a large-scale real-world multi-modal sensing and communication dataset," *IEEE Commun. Mag.*, 2023.

[158] J. Pegoraro, P. Saucedo, J. O. L. Jucht, M. Rossi, and J. Widmer, "Disc: a dataset for integrated sensing and communication in mmwave systems," 2022. [Online]. Available: https://dx.doi.org/10.21227/2gm7-9z72

[159] J. Luo *et al.*, "Sensiverse: A dataset for isac study," 2023.

[160] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, "Wideband spectrum sensing in cognitive radio networks," in *IEEE Int. Conf. Commun.*, 2008, pp. 901–906.

[161] S. Z. Gurbuz, H. D. Griffiths, A. Charlish, M. Rangaswamy, M. S. Greco, and K. Bell, "An overview of cognitive radar: Past, present, and future," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 34, no. 12, pp. 6–18, 2019.

[162] C. K. Kim, C. S. Yang, and C. G. Kang, "Adaptive listen-before-talk (LBT) scheme for LTE and Wi-Fi systems coexisting in unlicensed band," in *Proc. 13th IEEE Annual Consumer Commun. & Netw. Conf. (CCNC)*, 2016, pp. 589–594.

[163] A. Martone and M. Amin, "A view on radar and communication systems coexistence and dual functionality in the era of spectrum sensing," *Digit. Signal Process.*, vol. 119, 12 2021.

[164] X.-L. Zhu, Y.-A. Liu, W.-W. Weng, and D.-M. Yuan, "Channel sensing algorithm based on neural networks for cognitive wireless mesh networks," in *Proc. 4th Int. Conf. on Wirel. Commun., Netw. and Mobile Comput.*, 2008, pp. 1–4.

[165] S. Zheng, S. Chen, P. Qi, H. Zhou, and X. Yang, "Spectrum sensing based on deep learning classification for cognitive radios," in *China Commun.*, 2020, pp. 138–148.

[166] Y. Zeng and Y.-C. Liang, "Eigenvalue-based spectrum sensing algorithms for cognitive radio," *IEEE Trans. Commun.*, vol. 57, no. 6, pp. 1784–1793, 2009.

[167] C. P. Robinson, D. Uvaydov, S. D'Oro, and T. Melodia, "Deepsweep: Parallel and scalable spectrum sensing via convolutional neural networks," in *Proc. IEEE Int. Conf. on Machine Learning for Commun. and Netw. (ICMLCN)*, 2024, pp. 505–510.

[168] W. Zhang, Y. Wang, X. Chen, Z. Cai, and Z. Tian, "Spectrum transformer: An attention-based wideband spectrum detector," *Trans. Wirel. Commun.*, vol. 23, no. 9, pp. 12 343–12 353, 2024.

[169] D. Uvaydov, S. D'Oro, F. Restuccia, and T. Melodia, "Deepsense: Fast wideband spectrum sensing through real-time in-the-loop deep learning," in *Proc. IEEE Conf. on Computer Commun. (INFOCOM)*, 2021, pp. 1–10.

[170] K. Davaslioglu and Y. E. Sagduyu, "Generative adversarial learning for spectrum sensing," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, 2018, pp. 1–6.

[171] L. Cai, K. Cao, Y. Wu, and Y. Zhou, "Spectrum sensing based on spectrogram-aware CNN for cognitive radio network," *IEEE Wirel. Commun. Lett.*, vol. 11, no. 10, pp. 2135–2139, 2022.

[172] H. Wymeersch, J. Lien, and M. Z. Win, "Cooperative localization in wireless networks," *Proc. IEEE*, vol. 97, no. 2, pp. 427–450, 2009.

[173] W. Zhang, R. K. Mallik, and K. B. Letaief, "Cooperative spectrum sensing optimization in cognitive radio networks," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, 2008, pp. 3411–3415.

[174] D. Janu, K. Singh, S. Kumar, and S. Mandia, "Hierarchical cooperative LSTM-based spectrum sensing," *IEEE Commun. Lett.*, vol. 27, no. 3, pp. 866–870, 2023.

[175] N. A. Khalek and W. Hamouda, "Intelligent spectrum sensing: An unsupervised learning approach based on dimensionality reduction," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, 2022, pp. 171–176.

[176] W. Zhang, Y. Wang, X. Chen, L. Liu, and Z. Tian, "Collaborative learning based spectrum sensing under partial observations," *IEEE Trans. Cogn. Commun. Netw.*, pp. 1–1, 2024.

[177] Y.-C. Liang, Y. Zeng, E. C. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Trans. Wirel. Commun.*, vol. 7, no. 4, pp. 1326–1337, 2008.

[178] J. Aisbett, "Automatic modulation recognition using time domain parameters," *Signal Processing*, vol. 13, no. 3, pp. 323–328, 1987.

[179] J. Reichert, "Automatic classification of communication signals using higher order statistics," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP*, vol. 5, 1992, pp. 221–224 vol.5.

[180] O. Dobre, A. Abdi, and W. Su, "Survey of automatic modulation classification techniques: Classical approaches and new trends," *Communications, IET*, vol. 1, pp. 137 – 156, 2007.

[181] B. Kim, J. Kim, H. Chae, D. Yoon, and J. W. Choi, "Deep neural network-based automatic modulation classification technique," in *Proc. Int. Conf. ICT Converg. (ICTC)*, 2016, pp. 579–582.

[182] A. K. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bulettin of the Calcutta Mathematical Society*, no. 35, pp. 99–109, 1943.

[183] M. H. Shah and X. Dang, "Novel feature selection method using Bhattacharyya distance for neural networks based automatic modulation classification," *IEEE Signal Process. Lett.*, vol. 27, pp. 106–110, 2020.

[184] A. Fehske, J. Gaeddert, and J. Reed, "A new approach to signal classification using spectral correlation and neural networks," in *Proc. 1st IEEE Int. Symp. on New Frontiers in Dynamic Spectrum Access Netw.*, 2005, pp. 144–150.

[185] A. Swami and B. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 416–429, 2000.

[186] K. Triantafyllakis, M. Surligas, G. Vardakis, and S. Papadakis, "Phasma: An automatic modulation classification system based on random forest," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw. (DySPAN)*, 2017, pp. 1–3.

[187] H. Gang, L. Jiandong, and L. Donghua, "Study of modulation recognition based on HOCs and SVM," in *Proc. IEEE Veh. Technol. Conf. (VTC-Spring)*, vol. 2, 2004, pp. 898–902 Vol.2.

[188] V. Shatov, A. Vagollari, Y. Su, W. Gerstacker, M. Lübke, and N. Franchi, "Supervised learning of wireless signal waveforms in the era of radar-communications convergence," in *Proc. IFIP European Conf. on Netw. and Commun. (IFIP WMNC)*, 2025.

[189] A. P. Hermawan, R. R. Ginanjar, D.-S. Kim, and J.-M. Lee, "CNN-based automatic modulation classification for beyond 5G communications," *IEEE Commun. Lett.*, vol. 24, no. 5, pp. 1038–1041, 2020.

[190] Y. Shi, K. Davaslioglu, Y. E. Sagduyu, W. C. Headley, M. Fowler, and G. Green, "Deep learning for RF signal classification in unknown and dynamic spectrum environments," in *Proc. IEEE Int. Symp. on Dynamic Spectrum Access Netw. (DySPAN)*, 2019, pp. 1–10.

[191] T. K. Oikonomou *et al.*, "CNN-based automatic modulation classification under phase imperfections," *IEEE Wirel. Commun. Lett.*, vol. 13, no. 5, pp. 1508–1512, 2024.

[192] A. Kumar, M. S. Chaudhari, and S. Majhi, "Automatic modulation classification for OFDM systems using bi-stream and attention-based CNN-LSTM model," *IEEE Commun. Lett.*, pp. 1–5, 2024.

[193] R. Ding, F. Zhou, Q. Wu, C. Dong, Z. Han, and O. A. Dobre, "Data and knowledge dual-driven automatic modulation classification for 6G wireless communications," *IEEE Trans. Wirel. Commun.*, vol. 23, no. 5, pp. 4228–4242, 2024.

[194] F. Hlawatsch and G. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Sig. Process. Mag.*, vol. 9, no. 2, pp. 21–67, 1992.

[195] A. Vagollari, M. Hirschbeck, and W. Gerstacker, "An end-to-end deep learning framework for wideband signal recognition," *IEEE Access*, vol. 11, pp. 52 899–52 922, 2023.

[196] H.-K. Le, V.-P. Hoang, V.-S. Doan, M.-T. Hoang, and N. P. Dao, "Performance analysis of convolutional neural networks with different window functions for automatic modulation classification," in *Proc. Int. Conf. ICT Converg. (ICTC)*, 2022, pp. 153–157.

[197] W. Zhang, M. Feng, M. Krunz, and A. Hossein Yazdani Abyaneh, "Signal detection and classification in shared spectrum: A deep learning approach," in *Proc. IEEE Conf. on Comput. Commun.*, 2021, pp. 1–10.

[198] S. Dinh-Van, T. M. Hoang, B. B. Cebecioglu, D. S. Fowler, Y. K. Mo, and M. D. Higgins, "A defensive strategy against beam training attack in 5G mmwave networks for manufacturing," *IEEE Trans. Inf. Forensics Secur.*, vol. 18, pp. 2204–2217, 2023.

[199] J. L. Xu, W. Su, and M. Zhou, "Distributed automatic modulation classification with multiple sensors," *IEEE Sens. J.*, vol. 10, no. 11, pp. 1779–1785, 2010.

[200] O. Ozdemir, R. Li, and P. K. Varshney, "Hybrid maximum likelihood modulation classification using multiple radios," *IEEE Commun. Lett.*, vol. 17, no. 10, pp. 1889–1892, 2013.

[201] J. Zhang, D. Cabric, F. Wang, and Z. Zhong, "Cooperative modulation classification for multipath fading channels via expectation-maximization," *IEEE Trans. Wirel. Commun.*, vol. 16, no. 10, pp. 6698–6711, 2017.

[202] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, "Deep learning models for wireless signal classification with distributed low-cost spectrum sensors," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 3, pp. 433–445, 2018.

[203] Y. Wang, J. Wang, W. Zhang, J. Yang, and G. Gui, "Deep learning-based cooperative automatic modulation classification method for MIMO

systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4575–4579, 2020.

[204] Y. Wang *et al.*, "Distributed learning for automatic modulation classification in edge devices," *IEEE Wirel. Commun. Lett.*, vol. 9, no. 12, pp. 2177–2181, 2020.

[205] Y. Chen, J. He, W. Jiang, Y. Zhang, S. Huang, and Z. Feng, "Toward collaborative and channel-robust automatic modulation classification for OFDM signals," *IEEE Wirel. Commun. Lett.*, vol. 13, no. 11, pp. 3187–3191, 2024.

[206] A. Gao *et al.*, "Attention enhanced multi-agent reinforcement learning for cooperative spectrum sensing in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 73, no. 7, pp. 10 464–10 477, 2024.

[207] J. Fontaine, A. Shahid, R. Elsas, A. Seferagic, I. Moerman, and E. De Poorter, "Multi-band sub-GHz technology recognition on NVIDIA's Jetson Nano," in *Proc. IEEE 92nd Veh. Technol. Conf. (VTC2020-Fall)*, 2020, pp. 1–7.

[208] A. Vagollari, V. Schram, W. Wicke, M. Hirschbeck, and W. Gerstacker, "Joint detection and classification of RF signals using deep learning," in *Proc. IEEE 93rd Veh. Technol. Conf. (VTC2021-Spring)*, 2021, pp. 1–7.

[209] G. Kong, M. Jung, and V. Koivunen, "Waveform classification in radar-communications coexistence scenarios," in *Proc. 2020 IEEE Global Commun. Conf. (GLOBECOM)*, 2020, pp. 1–6.

[210] T. Huynh-The, V.-P. Hoang, J.-W. Kim, M.-T. Le, and M. Zeng, "Wavenet: Towards waveform classification in integrated radar-communication systems with improved accuracy and reduced complexity," *IEEE Internet Things J.*, vol. 11, no. 14, pp. 25 111–25 123, 2024.

[211] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *Proc. Int. Conf. Eng. Appl. Neural Netw.* Springer, 2016, pp. 213–226.

[212] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE J. Sel. Top. Signal Process.*, vol. 12, no. 1, pp. 168–179, 2018.

[213] V. Sathyanarayanan, P. Gerstoft, and A. E. Gamal, "RML22: Realistic dataset generation for wireless modulation classification," *IEEE Wirel. Commun.*, vol. 22, no. 11, pp. 7663–7675, 2023.

[214] N. West, T. O'Shea, and T. Roy, "A wideband signal recognition dataset," in *Proc. IEEE Workshop Signal Process. Adv. Wirel. Commun. (SPAWC)*, 2021, pp. 6–10.

[215] J. Wicht, U. Wetzker, and V. Jain, "Spectrogram Data Set for Deep-Learning-Based RF Frame Detection," *Data*, vol. 7, no. 12, 2022.

[216] C. Tassie *et al.*, "Detection of co-existing RF signals in CBRS using ML: Dataset and API-based collection testbed," *IEEE Commun. Mag.*, vol. 61, no. 9, pp. 82–88, 2023.

[217] J. Li, I.-T. Lu, and J. Lu, "Cramer-rao lower bound analysis of data fusion for fingerprinting localization in non-line-of-sight environments," *IEEE Access*, vol. 9, pp. 18 607–18 624, 2021.

[218] J. Bai *et al.*, "Machine learning-based near-field emitter location sensing via grouped hybrid analog and digital XL-MIMO receive array," *IEEE Trans. Cogn. Commun. Netw.*, pp. 1–1, 2025.

[219] C. Zhan, M. Ghaderibaneh, P. Sahu, and H. Gupta, "DeepMTL: Deep learning based multiple transmitter localization," in *Proc. IEEE 22nd Int. Symp. on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2021, pp. 41–50.

[220] A. Zubow, S. Bayhan, P. Gawłowicz, and F. Dressler, "DeepTxFinder: Multiple transmitter localization by deep learning in crowdsourced spectrum sensing," in *Proc. 29th Int. Conf. on Computer Communications and Networks (ICCCN)*, 2020, pp. 1–8.

[221] N. Singh, S. Choe, and R. Punmiya, "Machine learning based indoor localization using Wi-Fi RSSI fingerprints: An overview," *IEEE Access*, vol. 9, pp. 127 150–127 174, 2021.

[222] I. Bizon, Z. Li, A. Nimr, M. Chafii, and G. P. Fettweis, "Experimental performance of blind position estimation using deep learning," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2022, pp. 4553–4557.

[223] I. Bizon, A. Nimr, G. Fettweis, and M. Chafii, "Indoor positioning using correlation based signal analysis and convolutional neural networks," in *Proc. 19th Int. Symp. on Wirel. Commun. Syst. (ISWCS)*, 2024, pp. 1–6.

[224] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 12, pp. 5586–5609, 2022.

[225] L. Milosheski, M. Mohorčič, and C. Fortuna, "Spectrum sensing with deep clustering: Label-free radio access technology recognition," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 4746–4763, 2024.

[226] V. Shatov, N. Shanin, B. Perner, Y. Su, N. Franchi, and M. Lübke, "LSTM-DNN-based joint wireless signal waveform classification and

[227] blind transmitter localization," in *Proc. 19th Int. Symp. on Wirel. Commun. Systems (ISWCS)*, 2024.

[227] V. Shatov, N. Shanin, T. Veihelmann, B. Perner, N. Franchi, and M. Lübke, "Dual-task supervised learning for distributed spectrum monitoring applications," in *Proc. 16th Joint European Conf. on Netw. and Commun. & 6G Summit (EuCNC/6G Summit)*, 2025.

[228] T. Gong *et al.*, "A comparison of loss weighting strategies for multi task learning in deep neural networks," *IEEE Access*, vol. 7, pp. 141 627–141 632, 2019.

[229] C. Studer, S. Medjkouh, E. Gönültas, T. Goldstein, and O. Tirkkonen, "Channel charting: Locating users within the radio environment using channel state information," *IEEE Access*, 2018.

[230] P. Stephan, F. Euchner, and S. ten Brink, "Angle-delay profile-based and timestamp-aided dissimilarity metrics for channel charting," *IEEE Trans. Commun.*, vol. 72, no. 9, pp. 5611–5625, 2024.

[231] F. Euchner, P. Stephan, and S. ten Brink, "Uncertainty-aware dimensionality reduction for channel charting with geodesic loss," in *Asilomar Conf. on Signals, Systems, and Computers*, 2024.

[232] F. Euchner, M. Gauger, S. Dörner, and S. ten Brink, "A Distributed Massive MIMO Channel Sounder for "Big CSI Data"-driven Machine Learning," in *Proc. 25th Int. ITG Workshop on Smart Antennas (WSA)*, 2021.

[233] R. Hu and A. Singh, "UniT: Multimodal multitask learning with a unified transformer," in *IEEE/CVF Int. Conf. on Computer Vision (ICCV)*, 2021, pp. 1419–1429.

[234] N. Wang *et al.*, "Collaborative perception datasets for autonomous driving: A review," *IEEE Sens. J.*, vol. 25, no. 16, pp. 30 255–30 274, 2025.

[235] G. Charan *et al.*, "Towards real-world 6G drone communication: Position and camera aided beam prediction," in *Proc. IEEE Global Commun. Conf.*, 2022, pp. 2951–2956.

[236] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, 2020.

[237] A. Schuster, T. Reissland, and R. Weigel, "Hybrid sensing and communications receiver front-end for mobile communications applications," in *Proc. IEEE MTT-S Latin America Microw. Conf. (LAMC)*, 2023.

[238] G.-B. Wu *et al.*, "A universal metasurface antenna to manipulate all fundamental characteristics of electromagnetic waves," *Nat. Commun.*, no. 14, 2023.

[239] A. Jabbar *et al.*, "60 GHz programmable dynamic metasurface antenna (DMA) for next-generation communication, sensing, and imaging applications: From concept to prototype," *IEEE Open J. Antennas Propag.*, vol. 5, no. 3, pp. 705–726, 2024.

[240] C. Huang *et al.*, "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," *IEEE Wirel. Commun.*, vol. 27, no. 5, pp. 118–125, 2020.

[241] L. Zhu *et al.*, "A tutorial on movable antennas for wireless networks," *IEEE Commun. Surv. Tutor.*, vol. 28, pp. 3002–3054, 2026.

[242] X. Shao, R. Zhang, and R. Schober, "Exploiting six-dimensional movable antenna for wireless sensing," *IEEE Wirel. Commun. Lett.*, vol. 14, no. 2, pp. 265–269, 2025.

[243] Z. Yang *et al.*, "Pinching antennas: Principles, applications and challenges," *IEEE Wirel. Commun.*, pp. 1–10, 2025.

[244] M. Engelhardt *et al.*, "Accelerating innovation in 6G research: Real-time capable SDR system architecture for rapid prototyping," *IEEE Access*, vol. 12, pp. 118 718–118 732, 2024.

[245] B. Clerckx *et al.*, "Multiple access techniques for intelligent and multifunctional 6G: Tutorial, survey, and outlook," *Proc. IEEE*, pp. 1–48, 2024.

[246] K. L. Ketshabetswe, A. M. Zungeru, B. Mtengi, C. K. Lebekwe, and S. R. S. Prabaharan, "Data compression algorithms for wireless sensor networks: A review and comparison," *IEEE Access*, vol. 9, pp. 136 872–136 891, 2021.

[247] W. Yu *et al.*, "A survey on the edge computing for the internet of things," *IEEE Access*, vol. 6, pp. 6900–6919, 2018.

[248] M. Keshk, B. Turnbull, E. Sitnikova, D. Vatsalan, and N. Moustafa, "Privacy-preserving schemes for safeguarding heterogeneous data sources in cyber-physical systems," *IEEE Access*, vol. 9, pp. 55 077–55 097, 2021.

[249] A. Alwarafy, K. A. Al-Thelaya, M. Abdallah, J. Schneider, and M. Hamdi, "A survey on security and privacy issues in edge-computing-assisted internet of things," *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4004–4022, 2021.

[250] Y. Chu, P. D. Mitchell, and D. Grace, "Aloha and Q-learning based medium access control for wireless sensor networks," in *Int. Symp. Wirel. Commun. Syst. (ISWCS)*, 2012, pp. 511–515.

[251] M. P. Mota, A. Valcarce, J.-M. Gorce, and J. Hoydis, "The emergence of wireless MAC protocols with multi-agent reinforcement learning," in *Proc. 2021 IEEE Globecom Workshops*, 2021, pp. 1–6.

[252] N. Keshtiarast, O. Renaldi, and M. Petrova, "Wireless MAC protocol synthesis and optimization with multi-agent distributed reinforcement learning," *IEEE Netw. Lett.*, vol. 6, no. 4, pp. 242–246, 2024.

[253] E. Eldeeb and H. Alves, "Offline and distributional reinforcement learning for radio resource management," in *Proc. Joint European Conf. on Netw. and Commun. & 6G Summit (EuCNC/6G Summit)*, 2025, pp. 1–6.

[254] A. Bereyhi, A. Vagollari, S. Asaad, R. R. Müller, W. Gerstacker, and H. V. Poor, "Device scheduling in over-the-air federated learning via matching pursuit," *IEEE Trans. Signal Process.*, vol. 71, pp. 2188–2203, 2023.

[255] H. Wang, Z. Kaplan, D. Niu, and B. Li, "Optimizing federated learning on non-iid data with reinforcement learning," in *IEEE Conf. on Comput. Commun. (INFOCOM)*, 2020, pp. 1698–1707.

[256] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. Vincent Poor, "Federated learning for internet of things: A comprehensive survey," *IEEE Commun. Surv. Tutor.*, vol. 23, no. 3, pp. 1622–1658, 2021.

[257] A. Tusha and H. Arslan, "Interference burden in wireless communications: A comprehensive survey from PHY layer perspective," *IEEE Commun. Surv. Tutor.*, vol. 27, no. 4, pp. 2204–2246, 2025.

[258] M. H. Yılmaz and H. Arslan, "A survey: Spoofing attacks in physical layer security," in *Proc. IEEE 40th Local Computer Networks Conf. Workshops (LCN Workshops)*, 2015, pp. 812–817.

[259] M. Han *et al.*, "RF sensing security and malicious exploitation: A comprehensive survey," 2025. [Online]. Available: https://arxiv.org/abs/2504.10969

[260] G. Focarelli, S. Zanini, I. Palamà, G. Bianchi, and S. Bartoletti, "Positioning security in 5G and beyond: Model and detection of physical layer threats," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2025.

[261] L. Papangelo, M. Pistilli, S. Sciancalepore, G. Oligeri, G. Piro, and G. Boggia, "Adversarial machine learning for image-based radio frequency fingerprinting: Attacks and defenses," *IEEE Commun. Mag.*, vol. 62, no. 11, pp. 108–113, 2024.

[262] L. Huang, W. Pan, Y. Zhang, L. Qian, N. Gao, and Y. Wu, "Data augmentation for deep learning-based radio modulation classification," *IEEE Access*, vol. 8, pp. 1498–1506, 2020.

[263] Y. Meng, P. Qi, S. Zheng, Z. Cai, X. Zhou, and T. Jiang, "Adversarial attack and reliable defense based on frequency domain feature enhancement for automatic modulation classification," *IEEE Trans. Inf. Forensics Secur.*, vol. 20, pp. 3731–3744, 2025.

[264] Y. Zhao *et al.*, "Boosting robustness in automatic modulation recognition for wireless communications," *IEEE Trans. Cogn. Commun. Netw.*, vol. 11, no. 3, pp. 1635–1648, 2025.

[265] H. Bouzabia, T. N. Do, and G. Kaddoum, "Generative AI-empowered resilient adaptive ISAC against adversarial machine learning attacks," *IEEE Trans. Veh. Technol.*, vol. 74, no. 12, pp. 18 984–19 000, 2025.

[266] D. Adesina, C.-C. Hsieh, Y. E. Sagduyu, and L. Qian, "Adversarial machine learning in wireless communications using RF data: A review," *IEEE Commun. Surv. Tutor.*, vol. 25, no. 1, pp. 77–100, 2023.

[267] A. Shahid *et al.*, "Large-scale AI in telecom: Charting the roadmap for innovation, scalability, and enhanced digital experiences," 2025. [Online]. Available: https://arxiv.org/abs/2503.04184

[268] A. Aboulfotouh, E. Mohammed, and H. Abou-Zeid, "6G WavesFM: A foundation model for sensing, communication, and localization," *IEEE Open J. Commun. Soc.*, vol. 6, pp. 6792–6807, 2025.

[269] G. Pan, K. Huang, H. Chen, S. Zhang, C. Häger, and H. Wymeersch, "Large wireless localization model (LWLM): A foundation model for positioning in 6G networks," 2025. [Online]. Available: https://arxiv.org/abs/2505.10134

[270] M. Cheraghinia, E. De Poorter, J. Fontaine, M. Debbah, and A. Shahid, "A foundation model for wireless technology recognition and localization tasks," *IEEE Open J. Commun. Soc.*, vol. 6, pp. 9879–9896, 2025.