Physical Communication 75 (2026) 103036

Contents lists available at ScienceDirect

Physical Communication

journal homepage: www.elsevier.com/locate/phycom

Review article

Techniques for physical layer equalization and monitoring in radio-over-fiber: DSP and machine learning perspectives

César A. Montoya Ocampo *, Kevin D. Martinez Zapata, Jhon J. Granada Torres

Universidad de Antioquia, Facultad de Ingeniería. GITA lab, Calle 67 No 52-108, Medellín, Antioquia, 050004, Colombia

ARTICLE INFO

Keywords:
Digital signal processing
Equalization
Impairment mitigation
Machine learning
Neural network
Radio over fiber

ABSTRACT

Radio-over-fiber (RoF) systems are essential to the convergence of optical and wireless networks, serving as key enablers for future high-capacity broadband and 5G/6G applications. However, scaling capacity requires effective strategies for channel equalization and impairment mitigation to overcome complex nonlinearities, particularly at high frequencies. Traditional digital signal processing (DSP) methods are nearing their limits in dynamic scenarios. Thus, this paper addresses a significant lack in the literature by providing a comprehensive bibliometric and systematic analysis to objectively define the evolving roles of established DSP and emerging machine learning (ML) within this critical domain. By analyzing the co-evolution of key research themes, our approach confirms that the field is characterized by a structural coexistence between the two paradigms, rather than simple replacement. The primary results demonstrate that ML techniques, particularly deep learning (DL), establish functional superiority in compensating complex, high-order impairments in high-frequency bands. However, we identify a critical research gap in intelligent monitoring; while ML has high potential for multi-parameter estimation, it remains a niche interest often handled separately across optical and wireless layers. This drives the most pragmatic solution: hybrid DSP + ML architectures. These architectures balance DSP's computational efficiency, which sets a demanding benchmark for resource consumption, with ML's intelligence to achieve superior performance under strict cost, size, and power (C-S-P) constraints. The main finding is that this integration is transforming the RoF physical layer, shifting the focus from purely corrective equalization toward predictive, system-level control. The next critical challenges for commercial deployment are the successful implementation of real-time hardware and the standardization of public datasets.

1. Introduction

The growing global demand for higher bandwidth and lower latency connectivity is relentlessly pushing the limits of wireless access networks. To meet the requirements of beyond 5G and 6G paradigms, radio-over-fiber (RoF) technology has emerged as a foundational architecture, offering a unique blend of high capacity and low loss of optical fiber with the mobility and ubiquity of wireless access [1]. RoF systems achieve this by using a centralized approach: a single, high-speed optical fiber link is used to transport radio frequency (RF) signals from a central location to multiple distributed antenna units. This design centralizes the complex and power-intensive signal processing tasks (such as modulation, demodulation, and frequency conversion) at a central office or baseband unit (BBU) [2]. Consequently, the remote antenna units are simplified, becoming passive or semi-passive components that primarily convert the optical signal back to an RF signal for wireless transmission. This architecture offers several key advantages, including increased capacity, simplified and more cost-effective remote units, and enhanced energy efficiency [3]. Beyond this fundamental framework, recent research has introduced specialized architectural variations to meet the demands of next-generation networks. These include architectures for multi-service integration within a shared infrastructure [4], the use of multicore fiber (MCF) to enhance system capacity [5], and the application of analog RoF (ARoF) as a scalable fronthaul solution for 5G networks that leverages photonic integration [2]. However, the performance of RoF links is inherently susceptible to a multitude of linear and nonlinear impairments stemming from both the optical and RF domains, including chromatic dispersion (CD), laser phase noise (PN), amplifier spontaneous emission (ASE), and the nonlinear characteristics of Mach-Zehnder modulators (MZMs) [6].

Traditionally, the mitigation of these impairments has been performed using digital signal processing (DSP) techniques, which form a mature and powerful toolkit. Nevertheless, the growing complexity and scale of modern wireless networks, driven by paradigms such as

* Corresponding author.
E-mail address: cesara.montoya1@udea.edu.co (C.A. Montoya Ocampo).

https://doi.org/10.1016/j.phycom.2026.103036
Received 17 October 2025; Received in revised form 29 December 2025; Accepted 2 February 2026
Available online 3 February 2026

5G-Advanced and 6G, require more sophisticated approaches. Conventional DSP often struggles to model complex and coupled impairments, especially in dynamic link conditions, creating a performance ceiling for these traditional systems [7]. In this challenging landscape, machine learning (ML) has emerged as a promising alternative, offering a paradigm shift from reactive compensation to proactive, data-driven optimization [8]. ML approaches, including deep learning (DL) models, which can learn complex, nonlinear relationships directly from data, are candidates suited to address the intricate challenges of these evolving RoF systems. The incorporation of ML into the research field of RoF is not entirely a recent development. A noteworthy and groundbreaking instance dates back to 2010, when a study showcased one of the earliest successful applications of an ML technique, specifically the K-means clustering algorithm, for recovering the RF carrier phase in a digital coherent receiver for an optically phase-modulated RoF link [9]. This approach was crucial for achieving error-free performance at a high bit rate of 2.5 Gb/s over deployed fiber, effectively paving the way for considering data-driven solutions for more complex signal recovery problems in RoF systems.

Since this pioneering work [9], the potential of ML has been increasingly recognized within optical communications. This trend is expected to continue as networks evolve to support beyond 5G and 6G services, making ML a key enabler for technologies, such as RoF. To better situate our contribution within this expanding research field, we have reviewed previous surveys and review papers related to ML applications in fiber-wireless systems. The reviewing process has revealed three relevant works. The first one, published in 2020 [10], provides a comprehensive and application-oriented overview of ML techniques in RoF systems, encompassing both physical layer impairments and network-layer functions, such as traffic classification and dynamic bandwidth allocation. The second work, published in 2024 [11], adopts a narrower scope, focusing specifically on ML approaches to mitigate fiber nonlinearities and evaluate their impact on key physical layer metrics such as bit error rate (BER) and signal quality. Finally, the third one, published in 2020 [12], presents a survey on ML techniques for optical performance monitoring (OPM) and modulation format identification (MFI), highlighting how these methods can enhance system observability and adaptive control in optical networks, thus including RoF systems. Although the last-mentioned survey does not focus exclusively on RoF, its emphasis on impairment monitoring and signal classification provides a useful conceptual bridge to the challenges faced at the RoF physical layer.

While these existing reviews offer valuable, deep dives into specific technical applications of ML, a comprehensive, macro-level quantitative analysis of the field's evolution and intellectual structure is absent. More specifically, no single review provides a systematic synthesis of the findings at the physical layer, which is the primary domain where ML's ability to handle complex impairments and its performance-complexity trade-offs are most evident. To address this critical gap, this paper presents a systematic state-of-the-art review on the integration of ML in RoF systems, with an explicit emphasis on physical layer applications. Our review moves beyond a traditional narrative summary by employing a dual-method approach: a bibliometric analysis to quantitatively map the research landscape, evolution, and conceptual structure of the field, complemented by a traditional literature review to provide a qualitative and in-depth analysis of key technical contributions and insights.

The remainder of this paper is structured as follows: Section 2 details the methodology for the bibliometric analysis, presenting the quantitative results from our exploration of word frequency, thematic trends, and conceptual clusters. Section 3 provides a qualitative review of the existing research, with a specific focus on key application areas at the physical layer, including equalization and demodulation (both DSP and ML-based approaches), as well as monitoring techniques. Section 4 synthesizes the insights from both the bibliometric analysis and the literature review to provide a critical discussion of the field's current state and future directions. Finally, Section 5 concludes the paper by summarizing the key findings and outlining a clear path for future research.

2. Bibliometric analysis

This section presents the results of a bibliometric analysis conducted using the Bibliometrix R package and its web-based user interface Biblioshiny [13]. The objective of the analysis is to identify trends and developments in the use of ML and DSP techniques within RoF systems, with a specific emphasis on the physical layer. These applications include signal conditioning, equalization, system monitoring and estimation, as well as impairment mitigation. The subsequent section outlines the four-part structure of the advanced query used to gather this corpus.

1. The technological scenario of interest is a set of common abbreviations including "rof", its analog "arof", and digital "drof" forms, the more specific use cases in the RF domain, such as "microwave over fiber", "millimeter-wave over fiber", "mm-wave over fiber", as well as hyphenated terms such as "radio-over-fiber".
2. The application, which includes signal conditioning and estimation techniques, such as "equalization", "channel equalization", "impairment mitigation", "non-linearity compensation", "distortion compensation", "ICI mitigation" (ICI - inter-carrier interference), and "ICI suppression". Monitoring and classification tasks are covered through terms that include "noise estimation", "SNR estimation" (SNR - signal-to-noise ratio), "OSNR monitoring" (OSNR - optical SNR), "Q-factor estimation", "BER estimation", "signal quality monitoring", "link monitoring", "modulation format identification", "modulation classification", "modulation recognition", "signal classification"; multi-channel systems are also considered with terms such as "channel spacing estimation", "spectral overlap detection", and "guard band estimation". Furthermore, resource management concepts are captured using keywords, such as "resource allocation", "spectrum allocation", "dynamic spectrum management", "adaptive modulation", "adaptive resource allocation", "spectrum optimization", and "power allocation".
3. The employed method is limited to ML and DSP techniques, as well as their common variants. For instance, we use "machine learning", "machine-learning", and its abbreviation "ML", along with a family of models widely recognized as "artificial neural network". Other ML approaches are included through terms such as "support vector machine", "convolutional neural network", "recurrent neural network", "long short-term memory", "k-nearest neighbors", "decision trees", "random decision forests", "gradient boosting", and "fuzzy logic". Regarding the signal processing field, we have particularly included "digital signal processing", "signal processing" and "DSP".
4. A process of excluding works with poor relevance to this review was performed using the following selection criteria. An exclusion clause was incorporated into the query to improve precision by filtering out unrelated topics. Specifically, works mentioning "q-rung", "q-rung orthopair", "orthopair fuzzy", "q-ROF", or "q-ROF fuzzy", "reinforcement on federated", or "rotation forest" were excluded due to acronym collisions with "RoF". Additionally, terms such as "power over fiber" and "power-over-fiber" were removed to exclude works focused on energy transmission through optical fibers, which, while important in other domains, fall outside the scope of this review, which focuses on signal processing and system-level techniques.

The advanced search query was carefully constructed in four parts to focus specifically on ML applications within RoF systems, as well as DSP methods, in order to trace the evolution of methods in this domain. To ensure that the analysis was based on peer-reviewed, high-quality, and recent scientific contributions, only conference papers, journal articles, and books were considered. The documents were also limited to those published in English to ensure that they could be properly analyzed within the scope of this review.

The documents were also limited to those published between 2018 and 2025. This time window was selected based on an observed inflection point in publication activity: 2018 marks a noticeable increase in scientific output related to the application of ML techniques within RoF

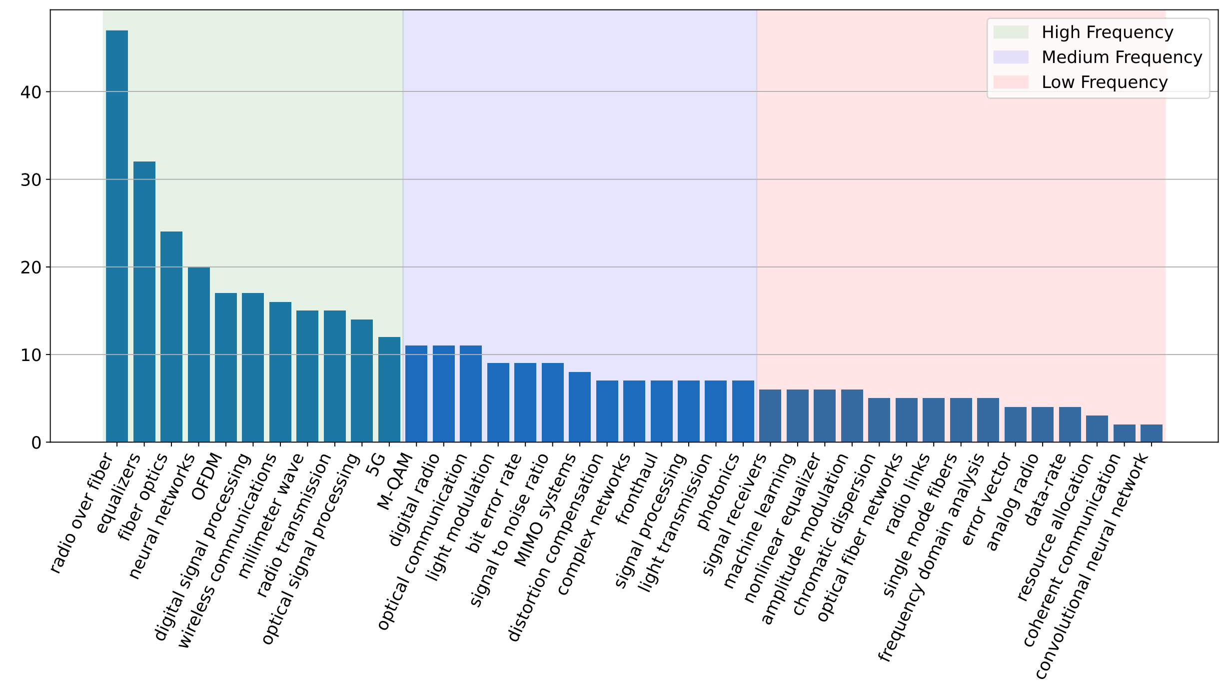

Fig. 1. Word frequency distribution for the corpus keywords. This visualization offers a global overview of the most recurrent terms across the dataset, providing a non-temporal snapshot of thematic prominence. The colored areas represent high, medium, and low-frequency zones, serving as a preliminary indication of which concepts are likely to appear more frequently in subsequent analyses.

systems. Thus, starting from this year ensures that the analysis captures the recent evolution and current dynamics in the research field while remaining representative of its most active period. A corpus of approximately 61 scientific publications was retrieved from the Scopus database using a query that included a filter to consider only those indexed in the most relevant subject areas for this study, namely Engineering, Physics and Astronomy, and Computer Science. Following this initial search, each document was meticulously reviewed to ensure its contribution was at the physical layer of the RoF system, thereby excluding papers that, while matching the initial query, focused on higher-level network functions not central to this analysis. This bibliometric analysis also includes an exploration of trending topics based on keyword co-occurrence, a thematic map to visualize relationships and the maturity of research themes, and a factorial analysis to cluster topics by conceptual similarity. These tools provide insights into the structure of current research, the prominence of specific themes, and how different concepts interrelate across the RoF literature, enhanced by ML and DSP techniques.

2.1. Word frequency and trend topics analysis

The frequency distribution shown in Fig. 1 provides a non-temporal overview of the most prominent concepts across the corpus, revealing the overall thematic relevance of keywords as measured by their frequency. In the high-frequency zone, which includes terms with more than ten occurrences, we find dominant topics such as “radio over fiber” (above 40), “equalizers” (slightly over 30), “fiber optics” (above 20), “neural networks” (20), as well as “digital signal processing”, “ofdm” (OFDM - orthogonal frequency division multiplexing), “wireless communications”, “millimeter wave”, “radio transmission”, “optical signal processing”, and “5g”, all of them ranging from 12 to slightly under 20 mentions. These terms are expected to recur in the thematic map and trend analyses due to their frequency and central role in the domain. In the mid-frequency range (around 7 to 11 appearances), we observe well-known concepts such as “m-qam” (QAM - quadrature amplitude modulation), “digital radio”, “optical communication”, “bit error rate”, “signal to noise ratio”, “mimo systems” (MIMO - multiple-input multiple-output), and “distortion compensation”, among others. These are likely to appear intermittently in more focused clusters or transitional topics. The low-frequency zone includes all terms with fewer than seven occurrences, such as “machine learning”, “chromatic dispersion”, “radio links”, “coherent communication”, and “convolutional neural network”. These may represent either emerging lines of inquiry or specialized concepts that, while relevant, have not yet achieved widespread presence in the field. With this initial analysis, we can see that neural networks, DSP, and equalizers are popular keywords in the corpus; this hints at them as central topics. Notably, extending the list to include more than the 40 most frequent keywords would mostly capture additional low-frequency terms with similar marginal appearances, offering limited additional insight. A more detailed examination of these temporal patterns is conducted in the following trend topic analysis, where the evolution of these terms over time is explored.

Fig. 2 presents a bubble chart of trend topics derived from the co-occurrence frequency of keywords in the curated bibliometric dataset. The x-axis represents the publication year, and the y-axis lists the most representative terms in the literature. Each bubble’s size and color reflect the frequency of the corresponding term, with red indicating higher frequency; the accompanying color bar quantifies this magnitude. Keywords are positioned along the time axis according to their peak relevance, while light gray horizontal lines indicate the interquartile range (IQR) of their temporal distribution. This visual representation allows us to trace the evolution of research priorities, identifying persistent themes, declining areas of interest, and newly emerging directions within the research field.

From the Fig. 2, we can see that optical communication has remained a foundational element in this field, along with other persistent concepts such as single-mode fibers, optical fiber networks, and digital radio. Well-established techniques, such as distortion compensation and amplitude modulation, also continue to appear, underscoring their long-standing relevance in RoF systems. A notable turning point

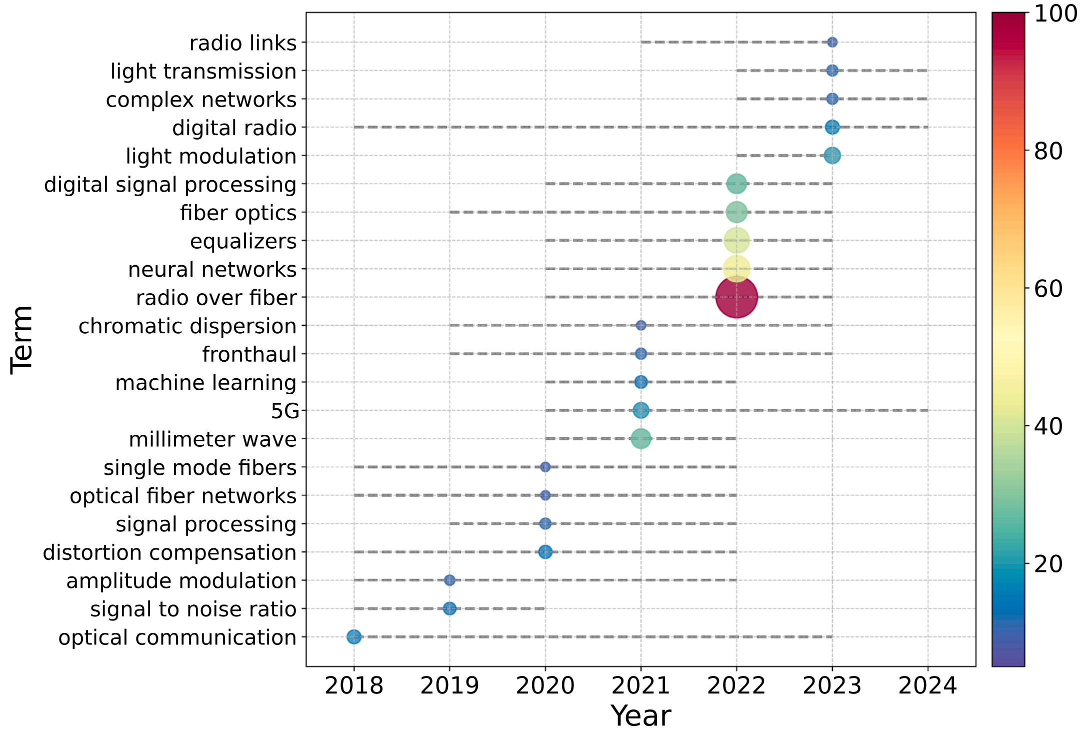

Fig. 2. Temporal distribution and prominence of research keywords related to ML and DSP in RoF systems. The bubble size and color indicate the relative frequency and intensity of each term, while their position along the time axis marks the period of peak relevance between 2018 and 2024.

occurred in 2021, when ML reached peak interest alongside 5G, millimeter wave (mmWave), fronthaul, and CD. The simultaneous rise of these terms suggests a growing concern with the limitations imposed by high-frequency operation, not only in the RF domain but also in the optical layer. In particular, the prominence of CD during this period reflects the fact that, as carrier frequencies increase and system bandwidths expand, dispersion effects in fiber links become increasingly critical, amplifying their impact on overall system performance. The simultaneous rise of these terms highlights a shift in focus toward end-to-end system considerations, where both optical and wireless impairments must be jointly addressed to enable next-generation services. This cluster of terms clearly reflects the evolution of mobile communication infrastructure, particularly the deployment and optimization of 5G networks. In 2022, the research field centers strongly on “radio over fiber”, the term with the highest overall frequency in the Fig. 2, an expected outcome, as it defines the thematic axis of the entire corpus. Its prominence following the surge in 5G interest suggests that RoF is being actively explored as a viable transport solution for next-generation mobile networks, with potential applications extending to future 6G systems. In parallel, the terms “equalizers” and “neural networks” also reached their peak in the same year, indicating a growing trend toward leveraging ML techniques for signal equalization. Similarly, the appearance of fiber optics and OFDM reinforces the field’s reliance on high-capacity optical transmission and spectrally efficient modulation techniques. Toward the end of the timeline, interest shifts to terms such as radio links, light transmission, complex networks, digital radio, and light modulation. The prevalence of light-related terms confirms the central role of fiber-optic systems in these studies, while the recurrence of radio links and digital radio points to specialized RF technologies gaining renewed attention. The mention of complex networks may relate to the increasing architectural complexity of RoF deployments, which span from heterogeneous fronthaul configurations to software-defined and virtualized network functions, which require adaptive and scalable orchestration strategies. These latest trends underscore the convergence of advanced photonic infrastructure with intelligent network design.

2.2. *Thematic analysis*

Thematic maps offer a macroscopic view of the structure and dynamics of a research field by analyzing co-word occurrences. They are built by computing two key metrics for each topic cluster: Callon centrality, which reflects the degree of interaction or relevance with other clusters (i.e., how integral the theme is to the field), and Callon density, which measures the internal cohesion or development of the cluster itself. Based on these two metrics, the map is divided into four quadrants, each representing a distinct type of research theme: motor, basic, niche, and emerging or declining. This dual-axis representation allows researchers to identify not only which themes are foundational or central but also which ones are emerging, declining, or niche. However, because this classification is derived solely from structural properties of the co-word network, additional temporal context can greatly enhance its interpretive power. In particular, by incorporating temporal trends from the trend topics analysis (see Fig. 2), it becomes possible to distinguish between emerging and declining themes within the same quadrant, as well as to validate the sustained influence or recency of more central topics. Fig. 3 presents the thematic map derived from our bibliometric corpus, including the most frequent words and summarizing the structure of the scientific discourse surrounding ML and DSP applications in RoF systems.

2.2.1. *Niche themes*

The niche themes quadrant, located in the upper-left quadrant of the thematic map (Fig. 3), highlights research areas with high internal development (density) but relatively low external connectivity (centrality). In bibliometric terms, these clusters often correspond to highly spe-

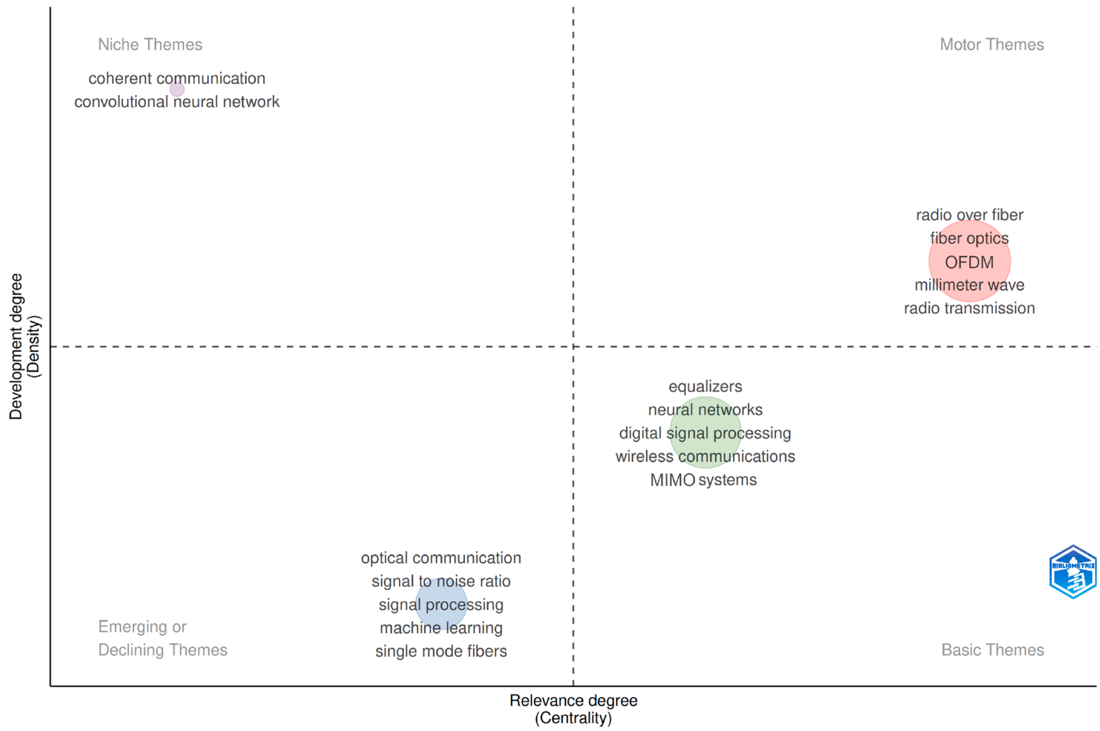

Fig. 3. Thematic map that categorizes research topics based on their degree of development (density) and relevance (centrality), forming four quadrants: motor themes (high centrality and density), basic themes (high centrality, low density), niche themes (low centrality, high density), and emerging or declining themes (low centrality and density). It highlights key clusters shaping the field.

cialized or self-contained topics that, despite their technical depth, interact less with the broader thematic structure of the field. It is essential to note that the boundaries between quadrants are not rigid; some clusters may lie near transition zones, and their categorization should be understood in relative rather than absolute terms. For instance, the lilac colored cluster represents a clear example of a niche theme. It is composed of just two terms: “coherent communication” and “convolutional neural networks”, both of which exhibit high internal coherence but low centrality in the thematic space. Their placement can be interpreted as follows: On the one hand, coherent communication is a well-established technique in optical transmission systems; yet, in the context of RoF, it is less prevalent. This is likely due to the widespread adoption of direct detection schemes in RoF links, which are simpler and more cost-effective, especially in access networks. As such, while coherent systems may appear in advanced research contexts, their influence across the broader RoF literature remains limited. On the other hand, convolutional neural networks (CNNs), though widely used in other ML domains, particularly for spatial data such as images, are not a natural fit for typical signal representations in RoF systems, which often rely on temporal or spectral characteristics. This reduces their citation and adoption within this niche, despite their methodological maturity. Their strong internal development is likely to reflect their consolidated role in the ML community, rather than their practical integration into signal-level processing for RoF systems.

2.2.2. Motor themes

The motor themes quadrant, located in the upper-right section of the thematic map (Fig. 3), typically represents topics that are both conceptually central and moderately developed. These themes are strongly connected to many other areas within the field and have accumulated a fair degree of internal coherence. However, in this particular case, the reddish cluster, while showing the highest centrality in the entire map, is positioned relatively close to the neutral axis of density. This suggests that, although its thematic influence is substantial, its internal cohesion is not absolutely high. As a result, its classification as a motor theme should be understood as relative, and its proximity to the boundary with basic themes indicates that it shares characteristics with both categories.

This cluster includes the terms “radio-over-fiber”, “fiber optics”, “ofdm”, “millimeter wave”, and “radio transmission”, which together represent a coherent technological backbone for hybrid optical-wireless systems. “radio-over-fiber” stands out not only because it was part of the search query but also due to its peak presence in 2022 and a sustained IQR from 2020 to 2023, as shown in Fig. 2, confirming its current centrality and supporting its classification as a motor theme. Similarly, “fiber optics” shows a comparable temporal footprint, with a slightly earlier onset (IQR from 2019), which suggests that while it remains relevant, its role is more foundational than dynamic, aligning more closely with the profile of a basic theme. “ofdm” does not appear in the trend topic analysis, which precludes direct temporal validation, yet its widespread use in high-capacity communication systems and its central position in the thematic map strongly suggest it functions as a motor concept. “millimeter wave” is temporally centered in 2021, with a presence from 2020 to 2022. Although this indicates active research during the early phase of 5G deployment, the limited range suggests it may be a recently emphasized basic technology rather than a currently dominant driver. “radio transmission” is absent from the trend topics plot, which, given its generality, is likely to reflect its function as a structural but thematically broad concept, characteristic of a basic theme. These temporal dynamics, together with the cluster’s centrality in the thematic map, reinforce the idea that it reflects core technologies that either drive or structurally support the RoF research landscape.

2.2.3. Basic themes

The basic themes quadrant, located in the lower-right region of the thematic map (Fig. 3), is typically associated with concepts that are highly central to the research field, but exhibit lower internal density. While centrality, in this context, reflects how strongly a theme is connected to other topics, a characteristic it shares with motor themes, its lower density is the key distinguishing factor. A low or moderate density, on the other hand, indicates that the thematic development within the cluster is either still in progress or not tightly focused, meaning the associated concepts are used across many contexts without forming a strongly self-contained body of research. It is important to clarify that the term “basic” does not imply simplicity or obsolescence. Rather, it refers to topics that are broadly relevant and widely cited, serving as essential building blocks in multiple research directions. In this particular map, the cluster (green one) identified within the basic themes quadrant is positioned very close to the origin, making it the closest to the center among all groups. This suggests that its centrality and density are both near the average values, reinforcing the idea that its role is neither strongly specialized nor marginal, but instead thematically balanced and widely applicable across the field.

This cluster contains the terms “equalizers”, “neural networks”, “digital signal processing”, “wireless communications”, and “mimo systems”. Its composition reflects the interaction between established signal processing methods and more recent ML approaches, situated within the broader context of wireless infrastructures that interface with optical backbones. As shown in Fig. 2, the term “equalizers” exhibits a temporal behavior very similar to that of “radio-over-fiber”, with a concentrated presence in recent years, suggesting that it remains to be a relevant and actively developed topic, consistent with its role as a motor theme. “neural networks” follows a nearly identical trajectory, with sustained activity across recent periods, reinforcing its status as an ML paradigm that has gained structural relevance in the design of modern compensation and estimation techniques. “digital signal processing” follows the same behavior in the trend topics figure, and as such, can be interpreted in the same light, as it underpins both traditional and ML-enhanced signal conditioning workflows. On the other hand, “wireless communications” does not appear in the trend analysis, probably due to its broad and general scope. This aligns with its characterization as a basic theme, foundational but not driving specific research directions. A similar interpretation applies to “mimo systems”, which, by definition, represent a standard architectural approach rather than an evolving research frontier. Overall, this cluster illustrates the thematic co-presence of algorithmic fundamentals and contemporary modeling techniques, forming a structural base that supports many of the innovations in the RoF landscape without necessarily being the source of conceptual disruption.

2.2.4. Emerging or declining themes

The emerging or declining themes quadrant, positioned in the lower-left region of the thematic map (see Fig. 3), is conceptually distinct because it does not represent a single type of topic, but rather a convergence of two opposing categories. What these themes share is their low centrality, meaning they are weakly connected to the broader thematic network, and low density, indicating that they are not yet well-structured internally. However, while both emerging and declining themes appear in this quadrant due to those shared properties, their temporal dynamics are fundamentally different. Emerging themes are typically new and still developing, which explains their low density and low network integration. In contrast, declining themes are losing relevance over time and no longer maintain strong internal cohesion. The position in this quadrant alone does not reveal which of these processes is at play. Therefore, proper interpretation requires external or complementary evidence, such as time-series keyword analysis or citation trends, in order to distinguish between growing novelty and diminishing importance.

In this map, the bluish colored cluster located in the lower-left quadrant includes the terms “optical communication”, “signal to noise ratio”, “signal processing”, “machine learning”, and “single mode fibers”. While this group initially appears thematically diverse, a closer examination using temporal trends (see Fig. 2) helps clarify the individual trajectories of each term and offers insight into their current bibliometric positioning. Starting with “optical communication”, its frequency peaked around 2018 and remains one of the most cited terms overall in the trend topics chart. This places it as a fundamental basic theme: although still present and clearly essential to the domain, it is a broad topic, rather than a specific one that identifies current research. Its bibliometric centrality is a direct result of it being so deeply embedded in the infrastructure of RoF systems that newer publications assume it as background knowledge rather than a focus of innovation. A similar case can be made for “single mode fibers”, which shares the same structural role in optical networks. These terms are appropriately classified as basic themes, as their infrastructural importance and declining prominence in recent metadata positions them closer to the central axis rather than the edge of research attention. “signal to noise ratio” exhibits a more sharply declining profile. According to the trend topics analysis, its peak occurred in 2019, with its IQR narrowing and ending around 2020. This suggests a reduced bibliometric presence in recent years. A plausible reason is that SNR, while still essential for performance evaluation, may have become a standardized metric that authors increasingly omit from keyword lists, especially when focusing on more novel aspects of system design or algorithmic innovation. “signal processing” occupies an intermediate position. Its decline is not as pronounced as SNR’s, but it lacks recent prominence. This could be due to its semantic ambiguity, spanning both optical and DSP, or because the field has shifted toward more specific subfields, such as ML-based processing or adaptive equalization, which authors may prefer to label more precisely. The presence of “machine learning” is particularly notable. Its frequency is centered around 2021. However, its IQR is short, the term is broadly scoped, and may be cited across disparate methodological approaches. This suggests that while ML is indeed a relevant and growing area within RoF systems, its bibliometric footprint may still be fragmented. The more specific and structurally integrated subtopics, such as “neural networks”, which appear in a separate cluster, are likely to absorb the more detailed attention in this phase of research development.

Altogether, this cluster reflects a transitional zone, where enduring technological pillars (such as “optical communication” and “single mode fibers”) coexist with shifting methodologies and metrics like “SNR” and “signal processing”. At the same time, “machine learning” hints at an evolving research direction that is still consolidating its place within the RoF literature. This conceptual tension makes the cluster a promising space for future exploration, where the integration of emerging techniques with legacy systems may define the next phase of innovation.

2.3. Factorial analysis

To explore the conceptual structure underlying the selected literature, a factorial analysis was conducted based on the 17 most frequent keywords appearing across the corpus. This analysis employed multiple correspondence analysis (MCA), a multivariate technique particularly well-suited for handling categorical data, such as co-occurrences of keywords in bibliographic metadata. MCA enables the reduction of dimensionality in complex datasets while preserving the geometric relationships between terms, facilitating the identification of thematic proximities and latent structures. Following the projection of terms into the reduced MCA space, a hierarchical clustering algorithm is applied to group keywords according to their spatial closeness in the factor space. The resulting dendrogram provides a visual representation of the similarity relationships among the terms, the x-axis ‘Distance’ corresponds to the dimensionless dissimilarity measure produced by the hierarchical clustering algorithm; larger values indicate weaker similarity between keyword clusters. As such, this axis in the dendrogram displays the progressive merging of terms into clusters based on their Euclidean distance within the MCA coordinates. The height at which branches merge reflects the dissimilarity between groups, with lower junctions indicating stronger semantic affinity. This tree-like structure thus reveals which concepts tend to co-occur more frequently within the literature and how thematic subgroups are hierarchically organized. In this case, the terms selected for this analysis correspond to key elements in the domain of RoF systems and their intersection with DSP and ML applications at the physical layer.

1. The orange cluster includes the terms “optical signal processing”, “light modulation”, “digital radio”, and “radio transmission”, forming a compact and technically coherent grouping. This cluster can be interpreted as representing a thematic axis centered on modulation and signal transport across hybrid optical-radio systems, where both optical and RF domains are treated symmetrically in terms of signal manipulation and propagation. For instance, the close association between “optical signal processing” and “light modulation” suggests a focus on the manipulation of optical carriers through digitally controlled schemes. This pairing is expected in the context of RoF systems, where modulating light efficiently and recovering the signal with minimal degradation is critical, especially under direct detection architectures. Optical signal processing refers to the suite of digital techniques applied to the optical domain, such as dispersion compensation, waveform shaping, or spectral filtering. Meanwhile, light modulation reflects the physical process that enables information to be encoded onto the optical carrier, and thus both terms tend to co-occur in studies discussing system design or performance analysis at the physical layer. On the other side of the cluster, the proximity between terms such as “digital radio” and “radio transmission” reflects a similarly tight relationship within the RF domain. “digital radio” typically refers to communication systems in which information is carried via digital modulation formats, such as QAM or phase-shift-keying (PSK), and this term is often used in RoF literature to emphasize the digitization of RF front-ends. “radio transmission”, in turn, emphasizes the physical emission and propagation of RF signals, and is a broader umbrella term that may include both analog and digital approaches. Their joint appearance indicates that discussions around digital modulation are still tightly coupled with radio propagation aspects in the literature, a characteristic especially relevant in RoF systems where the RF waveform must be preserved over the optical link with high fidelity.

   What makes this cluster particularly relevant is the symmetric structure of its two sub-pairs, optical and radio, both dealing with signal conditioning and transport, but on different physical layers. Their coexistence in the same cluster suggests that many studies approach RoF architectures from a unified signal-processing perspective, treating modulation, conversion, and propagation as an integrated chain that spans both optical and RF domains. This thematic coherence also highlights a research trend to address impairments and performance trade-offs jointly, considering the end-to-end behavior of the RoF link. As such, this cluster appears to represent a foundational layer of RoF system design, where modulation schemes and signal-processing strategies define the performance envelope across the hybrid medium.
2. The green cluster, composed solely of “signal to noise ratio” and “optical communication”, appears as a compact and isolated grouping within the dendrogram, positioned at a relatively deep level of the hierarchy. This suggests that these terms, while strongly associated with each other, are thematically distant from the rest of the field’s dominant clusters. Their tight linkage is conceptually sound, as SNR remains a fundamental performance metric in optical systems, particularly under direct detection, where linearity and power budget limitations heavily influence the quality of transmission. The peripheral location of this pair may reflect the fact that optical performance metrics are often treated independently from RF or ML-related components in RoF studies, indicating a possible thematic separation between traditional optical communication analysis and emerging RoF signal-processing approaches. This also suggests that while optical performance is acknowledged, it may not be the central focus in studies where RF signal processing or ML techniques dominate.
3. The red cluster, the largest and most centrally located in the dendrogram, encompasses a diverse but thematically cohesive set of terms that collectively reflect the dominant conceptual axis of the RoF research landscape. Notably, the structure of the dendrogram reveals two internally consistent subgroups, each capturing different but complementary facets of system-level research. The first subgroup includes “millimeter wave”, “equalizers”, “wireless telecommunication systems”, and “neural networks.” The pairing of “millimeter wave” with “equalizers” is expected, as mmWave signals are highly susceptible to multipath fading and frequency-selective distortions, which necessitate robust equalization strategies. The proximity of “wireless telecommunication systems” reinforces this connection, since mmWave transmission and equalization techniques are foundational to next-generation wireless standards, particularly in 5G and beyond. The inclusion of “neural networks” in this cluster may initially seem surprising, but its presence suggests an emerging methodological trend where ML techniques, particularly deep neural models, are being explored to replace or augment traditional equalizers, or to adaptively compensate for nonlinearities and hardware impairments in the wireless front-end. This alignment reflects a convergence between advanced signal processing and data-driven optimization in modern RoF systems. Meanwhile, the second subgroup consists of “bit error rate”, “radio over fiber”, “digital signal processing”, “ofdm”, and “fiber optics.” The close relationship between “bit error rate” and “radio over fiber” underscores the central role of BER as a standard performance metric in RoF experiments and simulations, where signal integrity across the optical link is typically assessed in terms of this metric rather than solely optical metrics. “digital signal processing” naturally links to both BER and RoF, as it encompasses the algorithmic toolbox used to mitigate signal impairments in the conversion chain. The inclusion of “ofdm” and “fiber optics” rounds out this subgroup with terms that situate DSP applications within specific modulation formats and transmission media. In particular, OFDM is relevant in RoF due to its spectral efficiency and resilience to CD, while “fiber optics” anchors the system physically in the optical domain. Altogether, this large cluster can be interpreted as a thematic backbone that integrates transmission technologies, signal processing methods, and performance evaluation metrics. In the light of this, the hierarchical proximity of the aforementioned terms reflects their frequent co-occurrence and methodological interdependence across the corpus. From a bibliometric perspective, the breadth and structure of this cluster highlight the maturity of RoF as a convergent research area, where optical, RF, and ML components are increasingly treated as parts of a unified design space.

2.4. Summary of the bibliometric analysis

From a bibliometric perspective, the analyzed publication corpus indicates that research on RoF systems is structured around a sustained coexistence between traditional DSP techniques and emerging ML approaches, rather than a disruptive paradigm shift. This observation is consistently supported by the thematic map and keyword co-occurrence analysis, where concepts such as equalizers, DSP, and neural networks appear jointly as central and foundational elements of the field. The temporal evolution of keywords further shows that this convergence became more pronounced around 2021, in parallel with the growing prominence of high-frequency RoF scenarios associated with 5G and mmWave technologies, suggesting a data-driven response to increasingly complex and coupled optical and RF impairments.

Importantly, the bibliometric findings do not point to ML as a fully independent solution space, but rather as a complementary methodological layer that enhances conventional signal processing frameworks. The factorial analysis highlights that terms related to mmWave, neural networks, and equalizers cluster closely, indicating that ML is predominantly investigated as an extension of equalization and impairment mitigation tasks, especially at higher carrier frequencies where analytical DSP models face practical limitations. In contrast, ML-based monitoring and parameter estimation appear less consolidated and remain sparsely represented in the literature, reinforcing the view that these applications have not yet matured into a distinct research direction. These quantitative insights provide the context and motivation for the following section, which examines representative DSP- and ML-based solutions in detail, emphasizing their system characteristics, practical implementations, performance trade-offs, and application scope.

3. Recent advances in equalization, impairment mitigation, and system monitoring with DSP and ML

To provide a clear, comprehensive overview of the methods and applications found in the literature, this section is divided into two parts. The first part examines DSP approaches for equalization and impairment mitigation, while the second does it for ML approaches. To facilitate quantitative comparison of these two specialized paradigms, Tables 1 and 2 provide a summary of their DSP and ML methods, respectively. The data collected for both tables includes: the reference of the paper (and its publication year), method employed, modulation format used, carrier frequency, data rate, system configuration (fiber distance and type, wireless distance and testing environment, "Exp." for experimental setups and "Sim." for simulation testbeds), the focused impairment of the proposed method, and the key metric obtained by the authors. When specific information was not reported in the original paper, the corresponding cell is marked as NR, and when the field is not relevant to the paper, N/A is used to specify it, as well as the reason in parentheses. Because some papers in the corpus are not focused on equalization or impairment mitigation but still contribute to ML approaches in RoF systems in the physical layer, we included these other papers related to monitoring and estimation of system parameters, which supports adaptive operation and performance optimization.

3.1. *DSP-based equalization and impairment mitigation*

Equalization and impairment mitigation in RoF systems have been extensively addressed through DSP, targeting both linear and nonlinear effects across different link configurations and frequency bands. These techniques are often tailored to the specific challenges posed by the carrier frequency, from low-frequency radio systems to high-frequency mmWave bands. At the sub-6 GHz and intermediate-frequency over fiber (IFoF) range, DSP techniques primarily focus on mitigating nonlinear distortions caused by components such as power amplifiers (PAs) and MZMs. For instance, in an experimental 20 km standard single-mode fiber (SSMF) RoF system carrying twelve 198-MHz f-OFDM subchannels with 64-QAM modulation (14.26 Gb/s aggregate, with no reported wireless link), the method compensated for distortions induced by the electro-absorption modulated laser and receiver photodiode (PD), achieving BER below the typical threshold in optical communications of 3.8×10^{-3} for hard decision forward error code (HD-FEC) with error vector magnitude (EVM) improvements against no equalization of $\sim$ 6 percentage points (pp) when the RF input power was -14.8 dBm [14]. In another work, an optimized magnitude selective affine digital predistortion (OMSA-DPD) method was introduced for the linearization of multiband RoF links carrying 5G new radio (NR) signals [15]. The scheme was validated experimentally using 5G NR signals at 2.14 GHz and 10 GHz, with channel bandwidths of 20 MHz and 50 MHz, but no explicit data rate. OMSA-DPD effectively mitigated PA nonlinearities and channel distortions by optimizing the affine function mapping on the signal envelope. The results showed a substantial reduction in adjacent channel leakage ratio (ACLR), reaching -46.1 dBc, and a marked improvement in EVM, achieving values as low as 1.9%, this represents an improvement of 1.7 dBc and 1.1 pp respectively, when compared with an MSA approach. Complexity was also addressed by calculating the number of coefficients and multiplications of the method and comparing it with MSA and a generalized memory polynomial (GMP), showing a lower complexity than both methods. Similarly, at 2.4 GHz, with no explicit data rate reported, a digital double sideband (DDSB) frequency translation technique combined with digital pre-distortion (DPD) was proposed to improve the linearity of RF MIMO transmission over fiber [16]. The method suppresses intermodulation distortion (IMD), which is particularly detrimental in MIMO systems. In a 16-QAM MIMO system, the approach achieved an EVM of 1–2% and a reduction of IMD products by 6-7 dB compared with the uncorrected case. By ensuring spectral purity and lowering nonlinear distortion, the DDSB-DPD technique enables more efficient use of the RF spectrum and enhances scalability for higher-order MIMO RoF deployments. Other studies have also applied blind nonlinear compensation methods that require no prior knowledge of the link. A distinctive example is a receiver-side decision-directed algorithm designed to monitor out-of-band nonlinear distortion components and iteratively suppress them using a steepest descent update. Unlike transmitter-side linearization techniques such as pre-distortion, feedback, or feedforward schemes (all of which require prior knowledge of the modulator curve, suffer from bandwidth limitations, or demand additional transmitters), the proposed approach operates blindly without calibration or extra hardware, while remaining suitable for broadband signals. In a more specialized IFoF application, a transmitter-side DSP scheme combining digital pre-coding and pre-compensation was applied to vector-modulated signals transmitted through electroabsorption modulators (EAMs). This approach directly mitigated the nonlinear phase distortion inherent to EAM-based modulation, enabling formats such as phase-balanced differential QPSK (PB-DQPSK), D8PSK, and D8-QAM. Experimental results showed clear improvements in signal quality against no equalization, with EVM as low as 3-4% and BER below HD-FEC threshold [17].

As carrier frequencies increase, particularly into the mmWave bands (28 GHz and above), DSP must address more severe impairments, including nonlinear distortion from electro-optic (EO) components, fiber dispersion, and burst noise in practical deployments. At 28 GHz, a cut-and-paste (CAP) based probabilistic shaping (PS) method combined with a Volterra nonlinear equalizer (NLE) was applied to a 64-QAM RoF link [18]. The system involved transmission over 10 km of SSMF with no wireless transmission, where the DSP chain effectively compensated the MZM nonlinearity and CD. The solution achieved a BER of less than 3.8×10^{-3}, effectively below the HD-FEC threshold and with a lower computational complexity in the form of less multiplications when compared against recursive least squares-decision feedback equalization (RLS-DFE), but 11.8% lower data rate. At 40 GHz, a cascaded multi-modulus algorithm (CMMA) was implemented as a blind equalizer for a pulse amplitude modulation (PAM)-4 signal transmitted over 25 km of SSMF and a 1 m wireless link [19]. The equalizer addressed general channel distortion in the RoF link, while its integration with interleaved Reed-Solomon (RS) coding specifically mitigated burst impulse noise. This combined DSP-FEC strategy enabled real-time performance with BER levels below 10^{-7}, when compared with no CMMA or RS the BER levels were around 10^{-3}, demonstrating a big upgrade in resilience against burst impulse noise and channel distortions. Researchers also estimated resource utilization and power consumption for both the transmitter and the receiver sides, they found that the power consumption was higher than that of the lasers, but was still relatively low for the entire system.

In the V-band (57-65 GHz), many DSP techniques have been investigated for both simulated and experimental RoF systems. At 60 GHz, a comparative simulation study evaluated different blind equalization algorithms, showing that the constant modulus algorithm (CMA) provided the best EVM performance for QPSK, while the least mean squares (LMS) algorithm performed better for 16-QAM, with an EVM improvement of 3.7 pp and 3 pp respectively when compared against no compensation, both effectively mitigating CD over up to 100 km of SSMF. Since this study primarily focused on optimizing the dispersive optical link, the distance of wireless transmission link was not reported [20]. Another study in this frequency band [21] experimentally demonstrated a frequency-domain Hammerstein NLE (FD-NLE) for mmWave RoF systems operating at 25 and 60 GHz, the proposed approach targets nonlinear distortions jointly introduced by the PA, MZM, and PD, and was validated using high-order OFDM modulation formats, including 64-QAM and 256-QAM, the experimental setup used 10 km of SMF followed by a short-range wireless hop of 1.2 m, the Hammerstein FD-NLE achieved a substantial reduction in signal distortion, lowering the EVM from 11.2% without equalization to 5.2%. In the same frequency range, an A-RoF architecture with intensity modulation/direct detection (IM/DD) was proposed for a cloud radio access network (C-RAN) fronthaul, where DSP was centralized at the BBU [22]. In this experimental setup, QPSK, 16-QAM, and 64-QAM signals were transmitted at data rates up to 24 Gb/s over 25 km of fiber, with the DSP enabling stable EVM values below 8.5% for 64-QAM, confirming the feasibility of centralized DSP in practical deployments. A simulation study explored OFDM demodulation aided by a combination of overlap frequency- and time-domain equalizers (OFDE + TDE) for 16-, 32-, and 64-QAM OFDM signals in a mmWave RoF scenario [23]. The scheme achieved a data rate of 100 Gb/s while reducing BER to 3.6×10^{-6} when dealing only with nonlinearities and 2×10^{-4} when including PN, showing that hybrid-domain DSP can effectively counteract nonlinearities and PN at these frequencies. Finally, in the upper end of the band, at 65 GHz, an experimental demonstration combined look-up table (LUT) predistortion with a CMMA blind equalizer for PAM-4 transmission [24]. The approach compensated for nonlinearities from EO components and PAs, successfully delivering a 13 Gb/s link with a BER of 2.23×10^{-3} over a 9 m wireless hop and no reported fiber length.

The W-band (92-94 GHz) pushes DSP to jointly address strong device nonlinearities, tight spectral occupancy, and frequency-synchronization errors. Three complementary solutions have been reported in this range. A photonics-aided demonstration employed probabilistically shaped (PS) OFDM with very high constellation orders (PS-128/PS-512-QAM) together with a hybrid time- and frequency-domain equalizer and a Volterra NLE [25]; this DSP chain enabled single-polarization transmission at a maximum aggregate raw rate of 208.4 Gb/s (PS-512-QAM configuration) while recovering signals with BER lower than the soft-decision FEC (SD-FEC) thresholds defined in ITU-T, namely 4.2×10^{-2} for a 25% overhead SD-FEC and 2.4×10^{-2} for a ∼20% overhead SD-FEC. For the lower modulation format order of truncated PS (TPS) 256-QAM and TPS-64-QAM, the lower SD-FEC limit was met. The key result in this research paper is the gain of achievable information rate (AIR) of 16.8 Gbit/s when using the proposed PS-NLE method. The Volterra stage mitigated device-induced nonlinear distortion, while the hybrid TDE/FDE equalizer corrected intersymbol interference (ISI), ICI and frequency synchronization errors. A second work demonstrated that such advanced DSP blocks can be implemented in hardware: a field programmable gate array (FPGA)-based real-time DSP framework incorporating a phase-folding / cumulative-transition frequency offset estimation (PF-CF-FOE) block, a decision-directed LMS (DDLMS) and a half symbol period CMMA (T/2-CMMA) equalizer delivered real-time equalization of 7.37-GBd QPSK signals [26], reducing DSP resource usage by approximately 68% compared with a Viterbi-Viterbi approach and achieving measured BERs on the order of 10^{-4} in the experimental link. Finally, in multi-antenna W-band links, blind MIMO equalization (CMA per stream) combined with maximal ratio combining (MRC) was shown experimentally to recover mmWave MIMO channels and deliver SNR gains up to 7.1 dB in a 2×2 setup [27], the experimental setup uses a 2 m wireless link, but doesn't specify any fiber link, confirming that classical blind algorithms remain effective when scaled to W-band MIMO RoF scenarios.

Finally, at the most extreme frequencies, such as 253 GHz, a Volterra-series nonlinear filter (VLNF) was deployed in a photonic-terahertz RoF system [28]. In this experimental setup, triangular waveforms of 500 MHz were transmitted over 10 km of single-mode fiber (SMF) and detected after a short-range wireless hop of 1-6 cm. The 253 GHz carrier was generated via optical heterodyning, using two closely spaced continuous wave (CW) lasers whose beating produced the sub-THz tone before modulation in the MZM. The Volterra-based DSP successfully compensated for both fiber dispersion and nonlinearities introduced by components such as the MZM, low-noise amplifier (LNA), PD, and Schottky-barrier diode (SBD). As a result, the system achieved signal-to-noise-plus-distortion ratio (SNDR) gains of 3.38 and 4.16 dB, corresponding to low- and high-received-power operating points, respectively. Despite a ∼25 dB difference in received power between these tests, the method demonstrated that advanced nonlinear equalization can extend the viability of photonic-terahertz RoF links even at sub-THz carrier frequencies.

Beyond these frequency-specific applications, DSP has also been applied to more general impairment-mitigation scenarios. One study examined a 16-channel dense wavelength division multiplexing RoF (DWDM-RoF) system carrying 16/32/64-QAM signals over 210-330 km of SMF with 50/100 GHz channel spacing, but without explicit note of the RF carrier frequency [29]. The receiver incorporated maximum likelihood sequence estimation (MLSE) combined with third-order nonlinearity compensation, supported by optical phase conjugation (OPC), Raman amplification, and dispersion-compensating modules (DCM) such as fiber Bragg grating (FBG). This DSP and optical compensation chain mitigated CD, Kerr nonlinearities, such as self-phase modulation (SPM), cross-phase modulation (XPM), and four-wave mixing (FWM), as well as residual ISI. For 16-QAM, the scheme enabled the aggregate transmission of 192 Gb/s across all 16 channels, achieving a BER of 10^{-9} and a Q-factor of 6, when using MLSE the maximum transmission distance was increased by 30 km for the 100 GHz and 60 km for the 50 GHz spacing scenarios, demonstrating an efficient mitigation of ISI and crosstalk. In a different context, a data-aided channel equalization method using Zadoff-Chu (ZC) training sequences was demonstrated in a RoF system designed to mimic the five-hundred-meter aperture spherical telescope (FAST) [30]. In this setup, training sequences were transmitted through approximately 60 m of G.657 fiber, representative of the 200-300 m spans in the telescope's cabling and no wireless link, it is important to note that this paper did not transmit digital information, nor did it use any modulation for the signal, the study was focused on addressing the phase drift and frequency impairments in fiber links for RF analog signals. The DSP algorithm interpolated the amplitude and phase response of the channel, successfully compensating for phase drift and frequency-dependent distortions in a dynamic environment. These examples underscore how DSP techniques extend beyond conventional fronthaul links to address time-varying channel impairments in both optical communication systems and large-scale scientific infrastructures.

3.2. ML-based equalization, impairment mitigation and monitoring

ML-based methods have recently gained popularity for equalization and impairment mitigation in RoF systems. Unlike conventional approaches, these methods utilize data-driven learning to approximate complex, nonlinear mappings between transmitted and received signals, making them adaptable to diverse channel conditions without requiring explicit modeling. A variety of architectures have been applied, including artificial neural networks (ANNs), CNNs, recurrent neural networks (RNNs), as well as specialized designs such as transfer-learning (TL)-assisted equalizers, U-Nets, and extreme learning machines (ELMs). Across different frequency bands and transmission scenarios, these methods have been reported to improve EVM, BER, and receiver sensitivity.

At lower microwave frequencies, ML-based equalizers have been tested in both S-band and X-band RoF links. In the S-band, a TL-assisted ANN-NLE (TL-ANN-NLE) was introduced to mitigate nonlinear impairments in an OFDM-64-QAM system with 200 MHz per user [31]. The setup consisted of a 2 km MCF link followed by a 0.5 m wireless transmission, targeting nonlinearities arising from the MZM and PD, as well as

IMD and cross-modulation distortion (XMD). The TL-ANN-NLE achieved a BER below the HD-FEC threshold with a 1.5 dB sensitivity improvement against the scenario without the proposed method, while also enabling the efficient reuse of trained models across users without requiring retraining from scratch, this reduced training times from 220 minutes to 36 minutes. In another work, a multilayer perceptron (MLP)-based ANN was designed for universal filtered multicarrier (UFMC) transmission with 4- and 16-QAM modulation [32]. This system, experimentally tested over 25 km SMF and simulated up to 100 km, followed by a 2 m wireless hop, achieved EVM of 0-10%, in contrast to 5-26% with conventional zero-forcing (ZF) equalization. Building upon this, a Fourier-transform network (FTnet)-based digital demodulator was developed for 16/64-QAM transmission [33]. In a 25 km SMF-only configuration, the FTnet reduced BER below the HD-FEC threshold, with performance gains of 6.9 dB compared to no equalization and 1 dB over LMS and conventional deep neural networks (DNNs), for higher-order 64-QAM, FTnet was the only demodulator to reach the HD-FEC limit, improving sensitivity by 3 dB over LMS. A subsequent study expanded this method to a 25 km SMF link followed by a 1-4 m wireless link [34], demonstrating BER below HD-FEC and receiver sensitivity down to -10 dBm. This data-driven demodulator outperformed LMS-, fully connected neural networks (FCNN), and Transformer-based baselines, and was the only method to reach the HD-FEC threshold under 64-QAM and long-range wireless conditions. Together, these works demonstrate that ANN- and FTnet-based equalizers effectively mitigate CD and nonlinearities in lower-frequency RoF systems, resulting in significant improvements in BER, EVM, and sensitivity compared to traditional DSP approaches. An ANN optimized through a genetic algorithm (ANN-GA) was proposed to suppress nonlinear distortion in dual-tone RoF transmission [35]. This study was performed entirely in simulation, and neither data rate nor optical/wireless link distances were reported. The work used a 15 GHz two-tone signal together with a 3 GHz crosstalk component to model nonlinearities from the laser, MZM, and PD. The ANN-GA achieved strong distortion mitigation, reducing XMD from -75.1 dBm to -114 dBm and IMD3 from -81.1 dBm to -108.3 dBm (~84 dB suppression). Compared with traditional digital post-compensation methods that rely on small-signal analytical models, the ANN-GA required no mathematical link modeling and generalized across different link conditions. Another work proposed a physics-informed neural equalization strategy for A-RoF fronthaul based on directly modulated lasers and direct detection (DML-DD) links [36]. The work addresses the dominant chirp-dispersion interaction inherent to DML-DD architectures, which manifests as composite second-order (CSO) and composite triple-beat (CTB) distortions, by embedding analytical distortion models directly into the neural network structure. The proposed tri-module multiplicative network (TriMNet) separates linear, CSO, and CTB compensation into dedicated first-, second-, and third-order modules, significantly reducing training data requirements compared to fully data-driven neural networks. Experimentally, the method was validated using a low-cost 16.8 GHz DML over 10 km and 15 km of SSMF, achieving CPRI-equivalent capacities of 1.02 Tb/s and 0.84 Tb/s, respectively, at an 8% EVM threshold. Compared with conventional Volterra-based feedforward equalizers (VFEs) and FCNNs, TriMNet reduced the required training symbols by 50% and training epochs by 62.5%, while maintaining comparable computational complexity.

At higher frequencies in the Ka-band, a NN post-equalizer cascaded with a least squares (LS) equalizer was applied to OFDM and single-carrier 64-QAM transmission [18]. The system operated at 28 GHz over 10 km of SSMF, where impairments such as nonlinearities from the dual-electrode MZM (DE-MZM), PD, and RF components, along with the high peak-to-average power ratio (PAPR) of OFDM signals, were dominant. The NN-LS cascade achieved a BER below the HD-FEC limit for both 6 Gb/s 64-QAM and 2 Gb/s OFDM 16-QAM, outperforming the RLS-DFE baseline and the second-order Volterra equalizer in BER and constellation quality. However, for OFDM 64-QAM, the NN failed to reach the FEC limit due to inaccurate LS channel estimation, highlighting the model's sensitivity to high-PAPR signals. A different approach in the same band introduced a NN joint equalizer for carrier frequency offset (CFO) and sampling clock offset (SCO) compensation in OFDM systems [37]. Tested with 16-QAM at 28 GHz and a 5 GHz intermediate frequency (IF), this method over a 25 km SMF link, followed by a 3 m wireless hop, improved frequency tolerance from 0.5/3.5 ppm to values exceeding 100 ppm, enabling reliable operation under severe synchronization errors and outperforming ZF-based compensation. Although no data-rate was reported, the feedforward NN reuses the training sequence, reducing complexity and latency. In the C-band, an ANN pre-equalizer was designed for high-order OFDM 1024-QAM transmission [38]. The experiment delivered 1.92 Tb/s over 500 m of SSMF, where impairments included ISI, nonlinear distortion, and high PAPR. The ANN pre-equalizer achieved an EVM of 2.38%, an SNR of 32.45 dB, and a receiver optical power (ROP) sensitivity of 1.8 dBm, standing as the only scheme meeting the 2.5% EVM target at 64 Gbaud. Although the complexity analysis was performed for post-equalization, the authors showed that a 3rd-order Volterra equalizer requires 4654 multiplications per symbol versus 4481 for an ANN, indicating that ANN-based schemes offer slightly lower but comparable complexity. Collectively, these works demonstrate that across Ku-, Ka-, and C-bands, ML methods significantly enhance the robustness of RoF systems against nonlinearities, synchronization errors, and PAPR-induced distortion while enabling capacities up to the terabit scale.

In the V-band, where the effects of CD are a critical aspect to address in these RoF systems [39], a complex-valued ANN-NLE with multi-level activation functions was introduced for single-carrier frequency-division multiple access (SC-FDMA) transmission using QPSK and 16-QAM formats [40]. The SC-FDMA signal had a bandwidth of 400 MHz and was transmitted on a 60 GHz carrier over a fiber-wireless link comprising 15 km SMF and a 0.8 m wireless hop. Nonlinearities from the MZM, EAM, and PD, as well as intra- and inter-band cross-modulation effects, were the main impairments. The ANN-NLE was evaluated in a multi-user A-RoF fronthaul and showed that joint multiuser equalization was essential to suppress inter-band XMD in the shared fiber channel. While BER fell below the HD-FEC threshold for intra-band operation, inter-band scenarios required joint equalization, with the ANN-NLE showing clear advantage at higher IF drive levels. In another work, an ELM equalizer was proposed for OFDM-based RoF links at 60 GHz [41]. This study, carried out through numerical simulations, modeled a QPSK-OFDM signal at 3.75 Gb/s over distances up to 1250 km of SMF. The ELM, particularly in its fully complex form, effectively mitigated CD, PN, ICI, and common phase error (CPE). Compared to pilot-aided equalization (PAE), the complex ELM achieved lower BER and nearly matched the performance of ideal CPE compensation while preserving the effective bit rate (no additional preamble was required). Its computational cost was also lower, requiring 56 ms per OFDM frame versus 64 ms for PAE. At the HD-FEC threshold, the method extended the laser linewidth tolerance from 0.48 MHz to 0.75 MHz at 14 dB SNR, outperforming conventional pilot-based correction. A different study applied an NN-NLE to a 3×3 MIMO RoF system using OFDM with 298 subcarriers at 60.5 GHz [42]. The link comprised 7 km SMF and up to 100 m wireless, where PA-induced nonlinearities with memory dominated. The NN-NLE raised SNR from 15.4 to 19.6 dB at high PA input power, boosted throughput by up to 30%, and extended the 100 Gb/s reach from ~40 m to 100 m, enabling up to 130 Gb/s transmission BER below the HD-FEC limit.

At higher frequencies in the W-band, recurrent neural network models have demonstrated strong nonlinear mitigation capability. A dual-gated recurrent unit (GRU)-based NLE was proposed for 64-QAM RoF transmission [43]. The link included 10 km of SMF and 1.2 m of wireless transmission, with impairments from the MZM, PD, LNA, mixer, and CD. The dual GRU equalizer reduced the BER below the SD-FEC threshold and improved receiver sensitivity by 2 dB relative to the CMMA and by 1 dB relative to third order Volterra equalization. In another work, a U-Net encoder-decoder neural equalizer was designed for 7.44 Gb/s PAM-4 transmission at 85 GHz [44]. In a purely wireless 300 m link, this method mitigated PA-induced nonlinearities, reducing BER to 2.1×10^{-3}, which was 42% lower than conventional CNN or FCNNs, while requiring 57% fewer floating-point operations (FLOPs). Finally, a two-lane demodulator (TLD) combined with the random oversampling (ROS) technique was also demonstrated for PS16-QAM at 88.5 GHz [45]. Using a 100 m SMF link followed by 4.6 km wireless transmission, this architecture compensates for nonlinearities from the MZM and PD as well as PS-QAM imbalance, achieving BER below HD-FEC while reducing computational complexity by 45.6%.

In the D-band, convolutional and complex-valued NNs have been applied to address nonlinearities and PN in RoF transmission. A two-dimensional convolutional NN (2D-CNN) was implemented as a post-equalizer for 25 Gbaud QPSK transmission at 128.75 GHz [46]. The system employed a 100 m SMF segment followed by a 4.6 km wireless link, where impairments included PN, as well as nonlinearities from the PD, PA, and MZM. The CNN-based equalizer reduced the BER to 5.3×10^{-3}, outperforming Volterra-, 1D-DNN-, and 1D-CNN-based equalizers. In another work, a complex-valued NN (CVNN) classifier combined with random oversampling (ROS) was proposed for PS-16QAM at 135 GHz [47]. Using 100 m of SMF and a 4.6 km wireless link, the CVNN-ROS equalizer mitigated nonlinearities from the MZM, PD, high-power amplifier (HPA), and mixers, as well as PS imbalance. The method reduced the BER below the SD-FEC threshold (1.4×10^{-2}), improved receiver sensitivity by 1 dB versus non-ROS TLD models, and required only 11k training samples. It also achieved a 27.7% reduction in computational complexity relative to the CVNN without ROS equalizer. A complementary solution introduced a joint DNN (J-DNN) equalizer combined with back-propagation (BP) and CMMA for PAM-8 transmission at 135 GHz [48]. Tested over a 10 km SMF link with a 3 m wireless hop, the J-DNN improved resilience against MZM, PD, and PA nonlinearities, achieving 60 Gb/s transmission with BER below the HD-FEC threshold and providing a 1 dB sensitivity gain over a standard DNN. These results highlight that in the D-band, CNN-, CVNN-, and DNN-based equalizers enable multi-gigabit RoF transmission with significantly improved BER and reduced complexity.

In the THz band, a complex-valued CNN (CV-CNN) was embedded into a TFDE scheme for OFDM RoF transmission [49]. The system operated at 340 GHz with 16-QAM at 16 Gbaud, transmitting 53.5 Gb/s over 20 km of SSMF, followed by 54.6 m wireless. Major impairments included CD, ISI, ICI, and nonlinearities from photoelectric conversion. The CV-CNN equalizer reduced BER below the SD-FEC threshold and improved receiver sensitivity by 0.9 dB over LMS-based TFDE, while outperforming a hybrid LMS-Volterra scheme in mitigating nonlinear distortion.

Beyond direct waveform equalization, several ML-based approaches reviewed in this work implicitly address tasks related to impairment estimation and system condition awareness, even when they are not explicitly formulated as monitoring solutions. For instance, one representative example addressed equalization in RoF systems without explicitly specifying the carrier frequency since the focus was on algorithmic performance rather than band-dependent characteristics. A multi-hidden-layer ELM (M-ELM) equalizer was proposed for QPSK-OFDM signals [50]. Designed to operate in real time using pilot subcarriers as training samples, the M-ELM3 (three-layer) architecture effectively compensated for PN, CD, and ICI without requiring additional preambles. Conducted entirely in simulation, this work showed that M-ELM3 achieved an increase of approximately 1.5 dB for 100 kHz and 1 MHz, and 3 dB for 2 MHz linewidths in the FEC limit when compared with PAE, R-ELM and two-real-input ELM (2RI-ELM) while approaching the effectiveness of complex ELM (C-ELM) but also preserving the full data rate. Although framed as an equalization problem, this approach demonstrates that ML models at the physical layer can learn channel characteristics and impairment signatures simultaneously.

In a similar algorithm-centric perspective, another work proposed an ML-based DPD scheme for A-RoF systems envisioned for 6G applications, focusing on linearization across cascaded optical and RF impairments rather than on band-specific operation [51]. The proposed approach employs an augmented real-valued time-delay neural network to jointly compensate for memoryless nonlinearities from the MZM, memory effects introduced by the PA, and linear CD. By explicitly incorporating a time-delay line at the network input, the model captures temporal correlations without relying on recurrent architectures, enabling effective mitigation of both dispersive and nonlinear distortions within a unified framework. The performance of the scheme was evaluated exclusively through numerical simulations using OFDM waveforms, demonstrating consistent improvements in EVM up to 7 pp compared to conventional Volterra NLE methods. However, the study does not explicitly report the RF carrier frequency nor the achieved transmission data rate, as the emphasis is placed on algorithmic robustness and adaptability across varying operating conditions rather than on specific system configurations.

From a broader perspective, these observations motivate considering ML-based monitoring not as an independent research line, but rather as a byproduct or extension of ML-assisted equalization and impairment mitigation in RoF systems. Unlike long-haul coherent optical links, where optical-domain observables are directly accessible, RoF systems typically rely on direct detection, and system impairments are mainly manifested in the electrical domain after photodetection. These include CD effects, noise accumulation, timing misalignment, and hardware-induced degradations. ML techniques applied at the physical layer can exploit this electrical-domain information to infer multiple system conditions concurrently, even when monitoring is not explicitly stated as the primary objective.

Recently, ML techniques originally motivated by signal processing tasks have also been shown to provide implicit awareness of signal conditions in RoF links. In this context, an autoencoder-based automatic modulation classification (AMC) scheme was experimentally validated for mmWave-over-fiber (MMWOF) transmission in [52]. The study considered six modulation formats (BPSK, QPSK, 8PSK, 16-QAM, 64-QAM, and 256-QAM) at a 28 GHz carrier and evaluated classification performance under different fiber lengths and OSNR levels. Experimental results, consistent with simulations, showed perfect classification at OSNR values above 15 dB for fiber spans up to 60 km. As OSNR decreased to 10 dB and 5 dB, accuracy degraded to approximately 87% and 82%, respectively. For fiber lengths beyond 70 km, dispersion-induced fading became the dominant impairment, leading to a marked reduction in classification accuracy, which dropped to around 40% at 100 km despite high OSNR. Nevertheless, for moderate spans such as 60 km, the classifier consistently maintained accuracies above 98% for OSNR values exceeding 10 dB, demonstrating that ML models can implicitly capture channel-degradation effects.

Related inference capabilities were also demonstrated in the classification of intercepted radar waveforms transmitted over RoF links [53]. In this work, low-probability-of-intercept (LPI) radar signals were intensity-modulated onto an optical carrier using an MZM, transmitted over SSMF, and recovered via direct detection. The ML-assisted task focused on waveform identification, including linear frequency modulation (LFM), frequency shift keying (FSK), and several polyphase codes. Feature extraction relied on Wigner distributions, Radon transforms, and ambiguity functions, followed by hierarchical decision-tree classification. To cope with fiber-induced CD and ASE noise, decision thresholds were adapted to the RoF channel conditions. Both simulation and experimental results over fiber spans up to 80 km showed average classification accuracies above 98% at OSNR values as low as 16 dB, with near-perfect performance at shorter distances or higher OSNR. These results indicate that even relatively lightweight, feature-based ML classifiers can provide robust signal awareness in RoF systems when impairment-aware adjustments are incorporated.

Beyond signal-level inference, ML has also been considered for equipment-oriented awareness in RoF deployments. In [54], a predictive health monitoring framework was conceptually proposed for RoF links carrying 64-QAM OFDM traffic. The architecture combines con-

Table 1
Compilation of DSP-enabled equalization techniques applied to RoF links, highlighting their modulation formats, operating bands, transmission rates, and the specific physical-layer impairments they address.

Ref.	Method	Mod. Format	Carrier Freq.	Data Rate	Distance	Impairment	Key Metric
Kim et al. [14]	Blind receiver-side nonlinear compensation (decision-directed)	64-QAM f-OFDM (12 × 198 MHz)	1.6 GHz IF	14.26 Gb/s	20 km SSMF; NR (Exp.)	Nonlinear waveform distortion (EAM, PD)	BER < HD-FEC; EVM -6 pp vs no EQ
Usman Hadi [15]	OMSA-DPD	256-QAM (5G NR signals)	2.14 GHz, 10 GHz	NR	10 km SSMF; no wireless (Exp.)	PA/modulator nonlinearities; ACPR	EVM: -1.1 pp; ACLR: -1.7 dBc vs MSA
Hussein and Liu [16]	DDSB frequency translation + DPD	16-QAM MIMO (3 channels)	2.4 GHz	NR	10 km SMF; no wireless (Exp.)	IMD (MZM nonlinearity)	EVM: 1-2%; IMD reduced by 6-7 dB vs no EQ
Mankong et al. [17]	Pre-coding + digital pre-compensation	PB-DQPSK, D8PSK, D8QAM	IFoF (EAM-based)	0.7-1 GSym/s	B2B fiber; no wireless (Exp.)	Nonlinear phase distortion from EAM	BER: $\sim 10^{-3}$; EVM: 3-4% vs no EQ
Song et al. [18]	CAP-based PS + Volterra NLE	64-QAM	28 GHz	6 Gb/s	10 km SSMF; no wireless (Exp.)	Nonlinear distortion from EO components + fiber dispersion	BER < HD-FEC; better complexity but 11.8% less data rate
Deng et al. [19]	CMMA blind equalization + interleaved RS coding	PAM-4	Q-band (40 GHz)	4.63 Gb/s	25 km SSMF; 1 m wireless (Exp.)	Channel distortion; burst impulse noise	BER $< 1 \times 10^{-7}$ vs no EQ 10^{-3}
Farooq and Miliou [20]	CMA, LMS, AMF	QPSK, 16-QAM OFDM	V-band (60 GHz)	NR	0-100 km SSMF; NR (Sim.)	CD	EVM gain: CMA 3.7 pp (QPSK); LMS 3 pp (16-QAM)
Li et al. [21]	Hammerstein FD-NLE	64-QAM, 256-QAM OFDM	25/60 GHz	6.8 Gb/s	10 km SMF; 1.2 m wireless (Exp.)	Nonlinear distortion from PA, MZM and PD	EVM: 5.2% vs 11.2% no EQ
Apostolopoulos et al. [22]	Centralized DSP-enabled A-RoF (C-RAN)	QPSK, 16/64-QAM	V-band (57-64 GHz)	up to 24 Gb/s	25 km SMF; 5 m wireless (Exp.)	Nonlinearities; EVM degradation (dispersion-induced fading, cross-modulation)	EVM: ≤ 8.5% (fiber-only); ≤ 13% (fiber-wireless)
Asif Khan et al. [23]	OFDE + TDE	16/32/64-QAM OFDM	57-61 GHz	100 Gb/s	20 km SMF; no wireless (Sim.)	Nonlinearities (SPM, XPM, FWM); PN	BER: 10^{-6} for NLs, 10^{-4} for PN
Zhou et al. [24]	LUT predistortion + CMMA	PAM-4	V-band (65 GHz)	13 Gb/s	NR; 9 m wireless (Exp.)	EO nonlinearities, PA distortion	BER: 2.23×10^{-3}
Wang et al. [25]	Hybrid TDE + FDE + Volterra NLE	PS-128QAM, PS-512QAM OFDM	W-band (92.5 GHz)	208 Gb/s	B2B fiber; 1 m wireless (Exp.)	Nonlinear distortion (device-induced) + ISI/ICI (frequency sync. error)	BER: < SD-FEC; AIR gain 16.8 Gbit/s vs no EQ
Zheng et al. [26]	FPGA DSP (PF-CT-FOE + T/2-CMA + DDLMS)	QPSK	W-band (92 GHz)	14.36 Gb/s	100 m SMF; 50 m wireless (Exp.)	Frequency/phase offset	BER: 1.2×10^{-4}; 68% less DSP vs Viterbi-Viterbi
Zhang et al. [27]	MIMO-CMA-MRC	QPSK	W-band (93.7 GHz)	64 Gb/s (32 GBaud × 2)	NR; 2 m wireless (Exp.)	Linear distortion	SNR gain: up to 7.1 dB (2 × 2 vs 1 × 1)
Shehata et al. [28]	Volterra NLE	16-QAM	253 GHz	N/A (triangular waveform)	10 km SSMF; 1-6 cm wireless (Exp.)	Dispersion + nonlinearities from EOM, PD, LNA, SBD	SNDR gain: 3.38-4.16 dB vs no EQ

Table 1
Continued.

Ref.	Method	Mod. Format	Carrier Freq.	Data Rate	Distance	Impairment	Key Metric
Sarkar et al. [29]	Combined OPC + Raman + MLSE	16/32/64-QAM DWDM-RoF (100/50 GHz spacing)	NR	192 Gb/s (16-QAM × 16 channels)	210-330 km SMF; no wireless link (Sim.)	Chromatic dispersion; Kerr nonlinearities (SPM, XPM, FWM); ISI	BER: 10^{-9}; Q-factor: 6; MLSE gain: +30-60 km vs no MLSE
Zheng et al. [30]	Data-aided channel equalization + ZC training	N/A (Analog RF signal)	70 MHz - 3 GHz	N/A (No digital data)	60 m G.657 fiber (representative of 200 m for telescope) + no wireless link (Exp.)	Phase drift; frequency-dependent amplitude/phase distortion	Corrected phase drift and frequency-dependent distortion

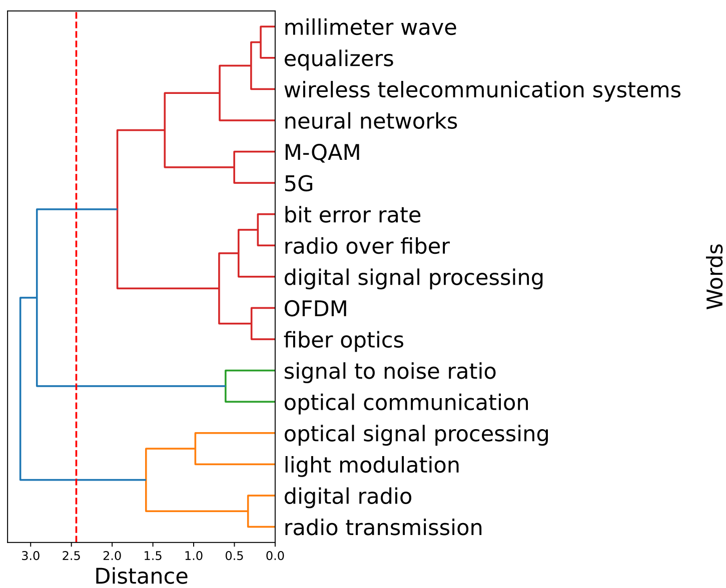

Fig. 4. Dendrogram derived from MCA using the 17 most frequent keywords in the corpus. The hierarchical clustering reveals the underlying thematic structure and proximities among terms.

ventional transmission metrics with auxiliary sensor data from optoelectronic components and suggests the use of ML techniques, including neural networks and fuzzy logic, for fault detection and prediction. While the scenario specifies a 12 km fiber link, no experimental validation, datasets, or quantitative performance metrics are reported. As such, this contribution primarily outlines a conceptual direction for integrating AI-assisted awareness at the device level, complementing ML-based equalization and impairment mitigation strategies, but leaving practical implementation and validation unaddressed.

4. Discussion on findings or insights

The systematic analysis of the literature, combined with the quantitative insights derived from the bibliometric mapping, reveals several pivotal themes regarding the role of DSP and ML in advancing RoF technologies. The following discussion synthesizes these findings, focusing on the structural coexistence of these two paradigms, the specialization of ML in high-frequency and nonlinear environments, the practical challenges of complexity and implementation efficiency, and the resulting shift toward system-level intelligence across both equalization and monitoring applications.

4.1. Structural coexistence and foundational role of DSP and ML

The synthesis of our quantitative and qualitative analyses establishes that the research field has matured into one of structural coexistence between traditional DSP and emerging ML techniques, rather than competition. This dynamic is evidenced by the persistent need for DSP as a mature, efficient baseline, which is simultaneously challenged and complemented by the specialized capabilities of ML. The continued relevance of DSP stems from its mathematical clarity and high efficiency in mitigating linear and specific nonlinear impairments.

For instance, DSP-based solutions demonstrate strong performance across various bands, achieving excellent quality metrics such as BER below the HD-FEC threshold and EVM under 2% in optimized systems. DSP's role can be categorized by the frequency band it serves. In low-to-mid frequency links, optimized DSP is indispensable for linearization and capacity management. This is exemplified by the OMSA-DPD, which achieved a low EVM of 1.9% for 5G NR signals [15], and the DDSB-DPD, which achieved an EVM of 1–2% for MIMO transmission [16]. Concurrently, DSP remains competitive in the high-frequency/mmWave bands, where it must confront severe dispersion and ISI. Examples include a hybrid TDE and FDE structure combined with a Volterra NLE to enable a 208 Gb/s transmission at 92.5 GHz [25], and the use of

Table 2
Overview of reported ML-based equalizers in RoF systems, organized by method, modulation format, carrier frequency, transmission distance, impairments mitigated, and resulting performance gains.

Ref.	Method	Mod. Format	Carrier Freq.	Data Rate	Distance	Impairment	Key Metric
Liu et al. [31]	TL-ANN-NLE	OFDM-64QAM (200 MHz/user)	S-band (3 GHz)	NR	2 km MCF; 0.5 m wireless (Exp.)	Nonlinearities (MZM, PD); IMD; XMD	BER < HD-FEC; +1.5 dB sensitivity vs no EQ
Liu et al. [32]	ANN (MLP, 1% training overhead)	UFMC (4/16-QAM)	X-band (10 GHz)	8 Gb/s (4-QAM UFMC)	25 km SMF (Exp.) up to 100 km SMF (Sim.); 2 m wireless	CD; MZM nonlinearity; high PAPR	BER: < HD-FEC; EVM 0–10% (ANN) vs 5-25% (ZF)
Zhu et al. [33]	FTnet-based digital demodulator	16/64-QAM	X-band (10 GHz)	8 Gb/s (16-QAM), 12 Gb/s (64-QAM)	25 km SMF; no wireless (Exp.)	CD; nonlinearities (MZM, PD)	BER < HD-FEC; +6.9 dB gain vs no EQ; +1 dB vs LMS/DNN; only method reaching HD-FEC in 64-QAM
Zhu et al. [34]	Data-driven FTnet digital demodulator	16/64-QAM	X-band (10 GHz)	8 Gb/s (16-QAM), 12 Gb/s (64-QAM)	25 km SMF; 1–4 m wireless (Exp.)	CD; nonlinearities (MZM, PD)	BER < HD-FEC; sensitivity -10 dBm; only method reaching FEC in 64-QAM/wireless
Yin et al. [35]	ANN-GA	Dual-tone RF signal	Ku-band (15 GHz)	NR	NR (Sim.)	XMD; IMD3 (laser, MZM, PD)	Distortion suppression: XMD -114 dBm; IMD3 -108 dBm
Zhang et al. [36]	TriMNet	64-QAM OFDM	16.8 GHz	1.02 Tb/s	10/15 km SSMF; no wireless (Exp.)	CSO, CTB and chirp dispersion	Bit rate: 1.02 Tb/s vs 0.9 Tb/s VFE; 50% less training data vs FCNN
Song et al. [18]	NN post-equalizer (cascaded with LS for OFDM)	16/64-QAM, OFDM	Ka-band (28 GHz)	6 Gb/s (64-QAM); 2 Gb/s (OFDM 16-QAM)	10 km SSMF; no wireless (Exp.)	Nonlinearities (DE-MZM, PD, RF); high PAPR (OFDM)	BER < HD-FEC; outperforms RLS-DFE/Volterra; fails for OFDM 64-QAM
Yan et al. [37]	NN joint CFO-SCO equalizer	16-QAM OFDM	Ka-band (28 GHz, 5 GHz IF)	NR	25 km SMF; 3 m wireless (Exp.)	CFO; SCO	BER < HD-FEC; tolerance 0.5/3.5 ppm → >100 ppm; outperforms ZF
Zhao et al. [38]	ANN pre-equalizer	OFDM 1024-QAM	C-band (32 GHz)	1.92 Tb/s	500 m SSMF; no wireless (Exp.)	ISI; nonlinear distortion; high PAPR	EVM 2.38%; SNR 32.45 dB; ROP sensitivity 1.8 dBm; only method meeting 2.5% EVM at 64 Gbaud; complexity comparable with 3rd-order Volterra
Liu et al. [40]	ANN-NLE (complex-valued, multi-level)	QPSK, 16-QAM SC-FDMA	V-band (60 GHz, 400 MHz BW)	266.7 Mb/s	15 km SMF; 0.8 m wireless (Exp.)	MZM, EAM, PD NLs; intra/inter-band XM	BER < HD-FEC (intra-band); > HD-FEC (inter-band; joint MU EQ required)
Zabala-Blanco et al. [41]	ELM equalizer (real/complex)	QPSK-OFDM	60 GHz	3.75 Gb/s	1250 km SMF; no wireless (Sim.)	PN; CD; ICI; CPE	BER < HD-FEC; linewidth tol. 0.48 to 0.75 MHz @ 14 dB SNR; C-ELM faster than PAE (56 vs 64 ms)
Wei et al. [42]	NN-NLE	OFDM (298 subcarriers)	V-band (60.5 GHz)	130 Gb/s	7 km SMF; up to 120 m wireless (Exp.)	PA NLs (memory effect)	SNR 15.4 to 19.6 dB vs no EQ; up to 30% higher data rate
Liu et al. [43]	Dual-GRU NLE	64-QAM	W-band (81 GHz)	60 Gb/s	10 km SMF; 1.2 m wireless (Exp.)	MZM/PD/LNA/mixer NLs; CD	BER < SD-FEC; +2 dB vs CMMA; +1 dB vs Volterra

Table 2
Continued.

Ref.	Method	Mod. Format	Carrier Freq.	Data Rate	Distance	Impairment	Key Metric
Zhang et al. [44]							
	U-Net encoder-decoder equalizer	PAM4	W-band (85 GHz)	7.44 Gb/s	B2B fiber; 300 m wireless (Exp.)	Nonlinearities (PA)	BER 2.1×10^{-3} 42% lower vs CNN/FCN; 57% fewer FLOPs
Xu et al. [45]							
	TLD equalizer + ROS	PS-16QAM	W-band (88.5 GHz)	10 Gbd/s	100 m SSMF; 4.6 km wireless (Exp.)	Nonlinearities (MZM, PD); imbalance (PS-QAM)	BER < HD-FEC; 45.6% less complexity with ROS
Jiang et al. [46]							
	2D-CNN post-equalizer	QPSK	D-band (128.75 GHz)	50 Gb/s	100 m SSMF; 4.6 km wireless (Exp.)	PN; PD, PA, MZM NLs	BER 5.3×10^{-3} < SD-FEC
Xie and Yu [47]							
	CVNN classifier + ROS	PS-16QAM	D-band (135 GHz)	11.4 Gb/s	100 m SSMF; 4.6 km wireless (Exp.)	Nonlinearities (MZM, PD, HPA, mixers); imbalance (PS-QAM)	BER < SD-FEC; 27.7% lower complexity against CVNN without ROS
Zhou et al. [48]							
	Joint DNN equalizer with BP + CMMA	PAM-8	D-band (135 GHz)	60 Gb/s	10 km SMF; 3 m wireless (Exp.)	nonlinearities (MZM, PD, PA)	BER < HD-FEC; 1 dB sensitivity gain vs DNN
Wang et al. [49]							
	Complex-valued Conv2D NN TFDE	16-QAM	THz band (340 GHz)	53.5 Gb/s	20 km SSMF; 54.6 m wireless (Exp.)	CD; ISI; ICI; nonlinearities (photoelectric)	BER < SD-FEC; +0.9 dB sensitivity vs LMS TFDE
Zabala-Blanco et al. [50]							
	M-ELM equalizer (OFDM pilots)	QPSK-OFDM	NR	NR	NR (Sim.)	PN; CD; ICI; CPE; AWGN	BER < HD-FEC; better than PAE, 2RI-ELM, R-ELM
Pereira et al. [51]							
	ARVTDNN	16-QAM OFDM	NR	NR	50–70 km SMF; NR (Sim.)	CD; nonlinearities (photoelectric)	EVM: 3% vs 10% no DPD

MIMO-CMA-MRC in the 93.7 GHz W-band to achieve a significant 7.1 dB SNR gain over a 1×1 system [27]. This collection of examples confirms that highly optimized DSP remains a robust baseline, essential for distortion compensation and capacity scaling across the entire RoF spectrum.

However, the bibliometric analysis provides the key structural insight into ML's foundational role. The themes of "equalizers", "neural networks" and "digital signal processing" cluster (see Fig. 4) together in the factorial analysis, and a simultaneous peak in activity around 2022 (see Fig. 2), confirming that all three are central to the current research landscape. Crucially, these concepts are situated within the "basic themes" quadrant of the thematic map (see Fig. 3), but are positioned notably close to the "motor themes" quadrant, the category reserved for the most active and driving themes in a field. This structural proximity signifies that ML is not merely a transient concept but a foundational component alongside established equalization methods, jointly guiding the current research trajectory.

This structural finding implies that ML's value is not in supplanting DSP, but in addressing the performance ceiling imposed by model-based DSP in increasingly complex, data-rich RoF scenarios. While DSP relies on explicit, pre-defined models of the channel and impairments, ML excels at implicit, data-driven feature extraction and pattern recognition. The coexistence suggests a complementary relationship where DSP handles the well-understood linear and low-order impairments, freeing ML to focus its significant processing power on the higher-order, coupled nonlinearities and dynamic channel variations that characterize next-generation RoF.

4.2. ML specialization in high-frequency and nonlinear regimes

The second major insight is the direct correlation between the adoption of ML and the increasing severity of impairments found in high-frequency, wide-bandwidth RoF systems. ML is no longer a mere alternative but the specialized tool for complex nonlinear and high-frequency environments where conventional model-based approach of DSP reaches its performance limit. This specialization is quantitatively evident when comparing frequency distributions: only 4 of the DSP-based methods are applied to carrier frequencies above or at the V-band (65 GHz) (see carrier frequency column in Table 1), whereas ML has 10 methods (compare with Table 2), this research interest in using ML for higher frequency systems is a result of the robustness of these methods that better compensate the impairments found in mmWave and above. To demonstrate the efficiency of these ML methods most of the research included a conventional DSP pipeline while also supplementing with specialized ML impairment mitigation blocks for better signal recovery, the studies results showed the efficiency of these methods by comparing with and without the ML blocks in the impairment mitigation pipeline. From the bibliometric perspective this specialization is driven by the demands of next-generation networks, a trend quantitatively confirmed by the bibliometric analysis. The thematic map (see Fig. 3) shows a tight cluster associating keywords like "millimeter wave" directly with "equalization" and "neural networks" themes. Moreover, the simultaneous rise in activity of "machine learning", "5G" and "millimeter wave" (see Fig. 2) demonstrates a clear technological shift: ML is a potential response to the extreme system impairments encountered as carrier frequencies move into the mmWave bands.

Technically, ML excels because its data-driven nature allows it to learn complex, coupled nonlinear relationships directly from the received signals, surpassing the performance ceiling of some of the presented conventional DSP techniques, such as Volterra or DPD, which rely on explicit channel models. Evidence from the review suggests that ML equalizers are most dominant in the upper frequency bands (see Table 2), where performance is most critical. While DSP has achieved success up to 93.7 GHz, the volume of ML-based papers in the V-band, W-band, and beyond is notably higher than that of DSP-based methods. For instance, ML-based techniques are reported at the highest carrier frequency investigated in the literature, 253 GHz [28], and are consistently applied to the most challenging formats, such as GRU equalizers at 81 GHz and U-Nets at 85 GHz for highly dispersive channels.

Furthermore, ML often shows superior robustness against mixed impairments. Where DSP can compensate for linear effects (such as CD) and moderate device nonlinearities, ML equalizers, including ANNs, GRUs, and CVNNs, outperform under the severe, coupled distortions inherent in high-speed, high-order modulation (see impairment column in Tables 1 and 2). A key example of ML based impairment mitigation, where the use of an ANN pre-equalizer in 1024-QAM transmission [38] was required to achieve the required EVM for a viable communication, other works also support this finding [8,55], although not specific to RoF systems, they're applicable as optical or RF standalone studies. This evidence solidifies the conclusion that ML is needed for pushing the boundaries of capacity and reach in the most challenging, higher modulation order, performance-critical, high-frequency RoF applications.

4.3. Complexity-performance trade-off and power consumption as a decisive factor

The adoption of ML in RoF systems is currently dictated by the trade-off between computational complexity and link performance. While ML excels in specialized regimes, DSP remains the benchmark for real-time implementation due to its superior efficiency in mitigating linear effects. This is exemplified by optimized FPGA-based DSP systems that achieve resource reductions of up to 68% [26], setting a demanding baseline that ML must yet overcome to be commercially viable. Despite this, recent efforts have focused in "lighter" ML architectures. Models such as ELMs [41,50], U-Nets [44], and CVNNs [47] have demonstrated complexity reductions between 40% and 50% while maintaining competitive BER. However, these advancements remain isolated cases rather than a systematic research trend, and many studies still frame their contributions around theoretical gains without evaluating hardware constraints.

A critical finding in this review is the methodological gap between algorithm validation and implementation feasibility. Complexity is predominantly quantified through proxy metrics–such as FLOPs, number of model parameters, or convergence time, rather than direct hardware measurements. Furthermore, power consumption remains largely unaddressed; while two DSP demonstrations explicitly report resource utilization [19,26], ML-based energy efficiency remains as an assumption rather than a direct measurement of experimental implementations. While not explicitly claimed, the common assumption of improved energy efficiency based solely on BER gains at lower received powers is misleading, as it ignores the computational overhead introduced by the ML processor itself. Consequently, without quantifying the net system-level energy balance, performance gains alone are insufficient to justify adoption. The lack of dedicated studies on real-time feasibility and embedded implementation continues to be the primary hurdle preventing ML equalizers from transitioning from laboratory prototypes to commercial RoF systems in constrained environments such as the mobile fronthaul.

4.4. The emergence of hybrid DSP + ML architectures

Recognizing the complexity-performance trade-off requires a practical solution that harnesses ML's superior mitigation capabilities without incurring excessive hardware costs. This need has driven the research community toward Hybrid DSP + ML architectures (here we refer to hybrid as systems that use non-indispensable DSP, such as Volterra or Viterbi-Viterbi NLEs and also ML methods, or learnable DSP, that combines ML BP within DSP blocks), which are emerging as the practical trade-off by achieving convergence between the two paradigms. These architectures operate on the principle of specialization: DSP handles the robust, computationally light tasks (such as linear equalization, frequency synchronization, or DPD), while the ML component is reserved for the computationally intensive task of learning and compensating for high-order nonlinear impairments.

This functional partitioning allows the resulting architecture to achieve a balanced performance profile, typically surpassing pure DSP methods while remaining lighter than unoptimized, purely ML-based DL models [56]. Multiple successful implementations illustrate this balance. For example, a NN cascaded with an LS equalizer was proposed for a 28 GHz system, significantly improving performance over traditional methods by distributing the equalization task [18]. Another notable example is the J-DNN combined with a CMMA in a 135 GHz system [48], which demonstrates the combination of a blind DSP equalizer with an intelligent NN structure to optimize convergence and accuracy. The complexity advantages of these hybrid methods are quantifiable. Architectures like the TLD + ROS structure [45] have been shown to reduce computational load by approximately 45.6% compared to monolithic, purely ML-based DL solutions, while retaining the performance edge required for high-capacity transmission. By intelligently limiting the scope and size of the ML portion and relying on efficient DSP for the foundational layer, these hybrid models successfully lower the resource demands necessary for real-time operation in FPGA and application specific integrated circuits (ASIC) systems.

This finding is supported by research in general optical communications. Several studies [57–61] use hybrid DSP + ML methods by turning the digital BP (DBP) step into learnable functions. This approach effectively handles linear and nonlinear effects in formats like PAM-4. These hybrid models outperform pure neural networks while keeping a low computational complexity comparable to traditional DSP. Ultimately, the development and increasing prevalence of these hybrid solutions strongly suggests the conclusion that the future of RoF equalization is defined by convergence, not replacement. These architectures represent the most pragmatic current approach to scaling capacity and operating at higher frequencies under tight constraints, proposing that the synergy between DSP's hardware maturity and ML's algorithmic intelligence is the key to technological viability.

4.5. Research gap in monitoring for autonomous RoF systems

The scarcity of ML-based monitoring reveals an imbalance in research priorities. While the community focuses heavily on ML as a corrective tool for equalization, using it to monitor system health remains a niche interest. Currently, intelligent monitoring is often handled as separate tasks for the optical and wireless layers rather than a unified RoF function. This suggests that ML is valued more for fixing signal errors than for the proactive, integrated management needed for "smart" networks. The few existing studies on modulation classification [52], radar signals [53], and equipment health [54] prove that ML is capable of high-level monitoring. However, these remain isolated efforts rather than a coordinated research trend. Because these works are fragmented, they do not yet provide a standard way to implement integrated monitoring solutions in real-world RoF hardware. Our bibliometric analysis confirms this gap, as monitoring keywords are absent from all trend maps and thematic analyses (Figure s 2, 3, and 4). This lack of visibility proves that the field is not yet mature. For RoF systems to reach commercial scale, research could benefit from a shift in focus–moving from just fixing signal errors to predicting them through integrated monitoring.

4.6. Thematic shift, system-level intelligence, and open challenges

The findings across the equalization domain collectively point toward a fundamental thematic shift in RoF research: moving from optimizing individual optical components to building intelligent, data-driven system-level control. The bibliometric analysis supports this, showing that foundational terms like "optical communication" and "signal to noise ratio" are classified as basic themes in the thematic map. This signals a maturity in the field where researchers are no longer focused on analyzing fundamental link physics but on the higher-level challenges of intelligent automation. The data-driven nature of ML enables the physical layer to be treated as a closed-loop control system, where impairment monitoring provides real-time system state (e.g., OSNR, skew), and equalization provides the adaptive actuation. This integration is essential for next-generation networks that require high resilience, automated fault detection, and advanced monitoring capabilities, including predictive health monitoring as recently suggested at the equipment level [54]. However, realizing this system-level intelligence requires addressing several critical open challenges that define the current research frontier:

1. **Complexity and real-time implementation:** the primary practical barrier remains the complexity and power efficiency trade-off. Despite advances in hybrid and lightweight architectures, the absence of extensive published results demonstrating real-time implementation of ML equalizers in hardware (FPGAs/ASICs) as well as power consumption evaluation confirms that this remains the dominant technological hurdle, an increased effort in this field should be considered in order to validate the real experimental implementation of these methods.
2. **Interpretable ML:** For ML to be trusted in complex systems, issues related to model black-box operation must be overcome. Future research must focus on developing interpretable ML models and exploring transfer learning strategies that allow models trained on one RoF link configuration to be effectively and reliably applied to another, reducing the time and data needed for training in diverse field environments.
3. **Data standardization and benchmarking:** Finally, the current landscape suffers from a lack of universally accessible, high-quality data. Because every study uses unique hardware, proprietary datasets, and distinct experimental setups, direct, apples-to-apples performance comparisons between different ML and DSP techniques are nearly impossible. To accelerate progress in the field, a critical recommendation is the establishment of publicly available, standardized RoF impairment datasets to allow for global benchmarking and the creation of reproducible, industry-relevant performance metrics. This crucial step is necessary to move research from isolated lab demonstrations toward field-ready, certifiable solutions.

5. Conclusions

To clarify the strategies for overcoming RoF physical layer limitations, this work performed a comprehensive bibliometric and systematic analysis of the literature, focusing on the specialized roles of traditional DSP and emerging ML techniques. The temporal analysis reveals that the research field is undergoing a transition from a long-established DSP framework toward the gradual incorporation of ML as a new processing paradigm. This transition is characterized by complementary coexistence: DSP remains the dominant and well-understood baseline for linear equalization and specific nonlinear equalization, while ML-based equalizers are increasingly adopted to address coupled nonlinear impairments and high-dimensional distortions that challenge conventional models. Particularly, neural networks have been gaining ground in tasks such as nonlinear impairment mitigation, monitoring, and parameter estimation. These converging research directions suggest that ML is becoming a foundational component of RoF research, guiding the field toward enhanced adaptability and system-level optimization.

Particularly, in this work, we found that an increased research interest in ML has enabled specialized functionalities in RoF systems that were largely inaccessible to traditional analytical or DSP methods. In the impairment mitigation domain, ML has emerged as a promising tool for addressing high-order nonlinearities and coupled distortions inherent in the highest-frequency bands (mmWave and above), where DSP-based models reach their performance ceiling. Additonally, ML based methods poses a new candidate in the domain of system observability, confirming technological viability for the application of automated multi-parameter monitoring as it is currently dominated by NN approaches. This capability provides an opportunity to enable the system to move beyond simple fault detection toward achieving multi-dimensional system visibility, a prerequisite for building the self-optimizing, adaptive networks of the future.

This pursuit of superior performance is inevitably tempered by the practical demands of deployment, leading to a significant issue regarding real-world viability. Since DSP has established a benchmark for cost, size and power (C-S-P) efficiency and hardware maturity, the superior mitigation of complex ML models must be integrated without incurring prohibitive computational and energy resources costs. This engineering challenge is being partially addressed with the hybrid DSP + ML (or learnable DSP) architectures. These systems are able to combine DSP's hardware maturity for computationally light, explainable, foundational tasks with ML's intelligence for complex nonlinear residuals. The successful implementation provided by these hybrid architectures can provide a viable path to scaling capacity and operating at higher frequencies.

The collective convergence of these specialized capabilities suggests a fundamental thematic shift in RoF research: moving away from the optimization of individual optical components toward building intelligent, closed-loop system-level control. The success of ML across both impairment compensation and detection allows the entire physical layer to be modeled as an adaptive, self-optimizing entity. This synthesis of findings suggests that the future development of RoF systems will be defined by their resilience, predictive health monitoring, and capacity for automated fault correction, all of which are enabled by the highly data-driven approach of ML.

The structural and functional insights detailed above delineate the significant findings of this work and illustrate the clear pathway toward fully autonomous RoF networks. However, realizing this transition remains constrained by several critical open challenges that define the immediate research agenda. The primary barrier remains the absence of extensive demonstrations of ML equalizers and monitors in real-time hardware (FPGAs/ASICs), proving that the complexity trade-off and limited power consumption evaluation is still a dominant hurdle to commercial viability according to the research gap found in our research. Furthermore, critical research must focus on enhancing the trustworthiness of ML models by developing interpretable architectures and applying transfer learning strategies to reduce retraining requirements in dynamic field environments. Finally, the field suffers from a lack of open-access datasets; virtually every study relies on proprietary, laboratory-specific datasets. To accelerate progress, the most crucial insight is the establishment of publicly available, standardized RoF impairment datasets and testing procedures. This critical step will enable global benchmarking, facilitate fair comparisons of model performance, and accelerate the deployment of certified solutions for industrial use.

CRediT authorship contribution statement

César A. Montoya Ocampo: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Kevin D. Martinez Zapata:** Writing – review & editing, Supervision, Resources, Methodology, Conceptualization; **Jhon J. Granada Torres:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the following projects from the University of Antioquia: (1) Prototipo de red óptica de acceso usando fuentes ópticas programables de grafeno y Machine Learning [2024-76210] and (2) SEMIN25-1-01 from the “Semillero SOFA”.

References

[1] A. Ng'oma, Radio-Over-Fibre Technology for Broadband Wireless Communication Systems, PhD Thesis, Technische Universiteit Eindhoven, 2005.

[2] S. Rommel, D. Dodane, E. Grivas, B. Cimoli, J. Bourderionnet, G. Feugnet, A. Morales, E. Pikasis, C. Roeloffzen, P. Van Dijk, M. Katsikis, K. Ntontin, D. Kritharidis, I. Spaleniak, P. Mitchell, M. Dubov, J.B. Carvalho, I. Tafur Monroy, et al., Towards a scaleable 5G fronthaul: analog radio-over-fiber and space division multiplexing, J. Lightwave Technol. 38 (19) (2020) 5412–5422. https://doi.org/10.1109/JLT.2020.3004416

[3] C. Lim, Y. Tian, C. Ranaweera, T.A. Nirmalathas, E. Wong, K.-L. Lee, Evolution of radio-over-fiber technology, J. Lightwave Technol. 37 (6) (2019) 1647–1656. https://doi.org/10.1109/JLT.2018.2876722

[4] M. Zhu, L. Zhang, J. Wang, L. Cheng, C. Liu, G.-K. Chang, et al., Radio-over-fiber access architecture for integrated broadband wireless services, J. Lightwave Technol. 31 (23) (2013) 3614–3620. https://doi.org/10.1109/JLT.2013.2286564

[5] L. Zhang, A. Udalcovs, R. Lin, O. Ozolins, X. Pang, L. Gan, R. Schatz, M. Tang, S. Fu, D. Liu, W. Tong, S. Popov, G. Jacobsen, W. Hu, S. Xiao, J. Chen, Toward terabit digital radio over fiber systems: architecture and key technologies, IEEE Commun. Mag. 57 (4) (2019) 131–137. https://doi.org/10.1109/MCOM.2019.1800426

[6] F. Keti, S.M.S. Atroshey, J.A. Hamadamin, A review of the impairments and challenges of radio over fiber technology and their mitigation strategies, in: 2022 International Conference on Computer Science and Software Engineering (CSASE), 2022, pp. 95–100. https://doi.org/10.1109/CSASE51777.2022.9759619

[7] A. Amari, O.A. Dobre, R. Venkatesan, O.S.S. Kumar, P. Ciblat, Y. Jaouen, et al., A survey on fiber nonlinearity compensation for 400 Gb/s and beyond optical communication systems, IEEE Commun. Surv. Tutor. 19 (4) (2017) 3097–3113. https://doi.org/10.1109/COMST.2017.2719958

[8] Q. Fan, G. Zhou, T. Gui, C. Lu, A.P.T. Lau, et al., Advancing theoretical understanding and practical performance of signal processing for nonlinear optical communications through machine learning, Nat. Commun. 11 (1) (2020) 3694. https://doi.org/10.1038/s41467-020-17516-7

[9] N.G. Gonzalez, D. Zibar, A. Caballero, I.T. Monroy, Experimental 2.5-gb/s QPSK WDM phase-modulated radio-over-fiber link with digital demodulation by a K - means algorithm, IEEE Photon. Technol. Lett. 22 (5) (2010) 335–337. https://doi.org/10.1109/LPT.2009.2039640

[10] J. He, J. Lee, S. Kandeepan, K. Wang, Machine learning techniques in radio-over-fiber systems and networks, Photonics 7 (4) (2020) 105. https://doi.org/10.3390/photonics7040105

[11] V. Jain, R. Bhatia, A survey on machine learning schemes for fiber nonlinearity mitigation in radio over fiber system, J. Opt. Commun. 45 (2024) s1157–s1163. https://doi.org/10.1515/joc-2022-0306

[12] W.S. Saif, M.A. Esmail, A.M. Ragheb, T.A. Alshawi, S.A. Alshebeili, Machine learning techniques for optical performance monitoring and modulation format identification: a survey, IEEE Commun. Surv. Tutor. 22 (4) (2020) 2839–2882. https://doi.org/10.1109/COMST.2020.3018494

[13] M. Aria, C. Cuccurullo, *bibliometrix*: an r-tool for comprehensive science mapping analysis, J. Informetr. 11 (4) (2017) 959–975. https://doi.org/10.1016/j.joi.2017.08.007

[14] B.G. Kim, S.H. Bae, H. Kim, Y.C. Chung, Blind compensation of nonlinear waveform distortions in radio-over-fiber system, in: 2018 Optical Fiber Communications Conference and Exposition (OFC), 2018, pp. 1–3.

[15] M. Usman Hadi, Towards optimization of 5G NR transport over fiber links performance in 5G multi-band networks: an OMSA model approach, Opt. Fiber Technol. 79 (2023) 103358. https://doi.org/10.1016/j.yofte.2023.103358

[16] S. Hussein, C.-P. Liu, Digital double-sideband frequency translation for transmission of RF MIMO signals over fiber, IEEE Microwave Wireless Technol. Lett. 33 (1) (2023) 110–113. https://doi.org/10.1109/LMWC.2022.3199907

[17] U. Mankong, P. Mekbungwan, K. Inagaki, T. Kawanishi, Vector modulation by laser and electroabsorption modulator with digital pre-coding and pre-compensation, J. Lightwave Technol. 36 (19) (2018) 4633–4639. https://doi.org/10.1109/JLT.2018.2846790

[18] T. Song, C. Lim, A. Nirmalathas, Performance analysis of a 28 GHz wideband analog radio-over-fiber fronthaul with channel nonlinearity compensation, J. Lightwave Technol. 42 (21) (2024) 7588–7595. https://doi.org/10.1109/JLT.2024.3416046

[19] R. Deng, J. Yu, J. He, Y. Wei, Experimental demonstration of a real-time PAM-4 q-band RoF system based on CMMA equalization and interleaved RS code, Opt. Fiber Technol. 42 (2018) 156–161. https://doi.org/10.1016/j.yofte.2018.03.006

[20] U. Farooq, A. Miliou, Assessment of different channel equalization algorithms for a converged OFDM-based 5G mm-wave a-RoF system at 60 GHz, Appl. Sci. 12 (3) (2022) 1511. https://doi.org/10.3390/app12031511

[21] P. Li, L. Zhu, H. Wang, W. Pan, N. Zhong, W. Bai, L. Yan, X. Zou, et al., One-shot blind frequency-domain nonlinear equalization for millimeter-wave fiber-wireless IFof mobile fronthaul system, J. Lightwave Technol. 41 (14) (2023) 4624–4634. https://doi.org/10.1109/JLT.2023.3243930

[22] D. Apostolopoulos, G. Giannoulis, N. Argyris, N. Iliadis, K. Kanta, H. Avramopoulos, Analog radio-over-fiber solutions in support of 5G, in: 2018 International Conference on Optical Network Design and Modeling (ONDM), 2018, pp. 266–271. https://doi.org/10.23919/ONDM.2018.8396143

[23] M.K. Asif Khan, F. Ali, M. Irfan, F. Muhammad, F. Althobiani, A. Ali, S. Khan, S. Rahman, G. Perun, A. Glowacz, Mitigation of phase noise and nonlinearities for high capacity radio-over-fiber links, Electronics 10 (3) (2021) 345. https://doi.org/10.3390/electronics10030345

[24] W. Zhou, P. Gou, K. Wang, M. Kong, X. Li, L. Zhao, Z. Zhu, J. Yu, PAM-4 Wireless transmission based on look-up-table pre-distortion and CMMA equalization at v-band, in: Optical Fiber Communication Conference (2018), Paper Th2a.60, Optica Publishing Group, 2018, p. Th2A.60. https://doi.org/10.1364/OFC.2018.Th2A.60

[25] K. Wang, L. Zhao, J. Yu, 200 Gbit/s photonics-aided MMW PS-OFDM signals transmission at w-band enabled by hybrid time-frequency domain equalization, J. Lightwave Technol. 39 (10) (2021) 3137–3144. https://doi.org/10.1109/JLT.2021.3062387

[26] T. Zheng, K. Wang, S. Hu, X. Yang, L. Zhang, C. Wang, B. Sang, C. Bian, J. Yu, Fpga-implemented real-time DSP framework enabling experimental 50 m photonics-aided millimeter-wave links with enhanced frequency offset compensation, Opt. Laser Technol. 191 (2025) 113287. https://doi.org/10.1016/j.optlastec.2025.113287

[27] Q. Zhang, J. Yu, X. Zhao, W. Li, M. Zhu, J. Zhang, Y. Wei, K. Wang, W. Zhou, B. Liu, F. Zhao, J. Yu, 2 × 2 MIMO W-band millimeter wireless signal delivery employing CMA and MRC technology, J. Lightwave Technol. 42 (13) (2024) 4485–4492. https://doi.org/10.1109/JLT.2024.3372408

[28] M. Shehata, K. Wang, W. Withayachumnankul, Mitigating the nonlinearity of radio-over-fiber terahertz systems, in: 2022 47th International Conference on Infrared, Millimeter and Terahertz Waves (IRMMW-THz), 2022, pp. 1–2. https://doi.org/10.1109/IRMMW-THz50927.2022.9895862

[29] N. Sarkar, B. Pal, B. Dutta, N. Sarkar, T.K. Chanda, A.S. Patra, Optimization of QAM modulated 16-channel DWDM-RoF system employing 3rd order nonlinearity compensation and MLSE equalization to transmit 192 Gbps data over long range through fiber, J. Opt. (2025) 1–11. https://doi.org/10.1007/s12596-025-02804-6

[30] Z. Zheng, Z. Su, X. Gao, G. Gao, J. Ma, J. Yang, C. Li, S. Huang, H. Liu, Data-aided channel equalization scheme for FAST radio over fiber transmission system, Opt. Express 29 (15) (2021) 24525–24535. Publisher: Optica Publishing Group, https://doi.org/10.1364/OE.433320

[31] E. Liu, Z. Yu, Y. Song, K. Sun, H. Huang, F. Yin, Y. Zhou, K. Xu, TL-ANN based nonlinear equalization for multi-users radio over fiber system, J. Lightwave Technol. 41 (5) (2023) 1399–1405. https://doi.org/10.1109/JLT.2022.3221166

[32] J. Liu, X. Zou, W. Bai, Performance enhancement of UFMC based radio over fiber system using ANN equalizer, in: Asia Communications and Photonics Conference (ACPC) 2019 (2019), Paper M4aA.331, Optica Publishing Group, 2019, p. M4A.331. https://opg.optica.org/abstract.cfm?uri = ACPC-2019-M4A.331.

[33] Y. Zhu, J. Ye, L. Yan, W. Pan, X. Zou, FTnet-Based digital demodulator for radio over fiber transmission, in: 2022 IEEE International Topical Meeting on Microwave Photonics (MWP), 2022, pp. 1–4. https://doi.org/10.1109/MWP54208.2022.9997709

[34] Y. Zhu, J. Ye, L. Yan, T. Zhou, X. Yu, P. Li, X. Zou, W. Pan, A Data-Driven Digital Demodulator Based on Deep Learning for Radio Over Fiber Transmission System, 2023, Issue: 23 Pages: 7192–7200 Publication Title: Journal of Lightwave Technology Volume: 41. https://doi.org/10.1109/JLT.2023.3282791

[35] Y. Yin, W. Yang, S. Xie, B. Yun, High-efficiency mitigation of nonlinear distortion in microwave photonics link assisted by artificial neural network, Wireless Commun. Mob. Comput. 2022 (2022) 1–8. https://doi.org/10.1155/2022/6327479

[36] Y. Zhang, Y. Zhu, L. Man, Q. Zhuge, W. Hu, et al., Rapid convergence and low data dependency: physics-informed Tri-module multiplicative network for DML-DD link in analog RoF fronthaul, J. Lightwave Technol. 43 (10) (2025) 4616–4623. https://doi.org/10.1109/JLT.2025.3538637

[37] J.-H. Yan, J.-S. Tseng, S.-C. Tsai, K.-M. Feng, A neural network based joint CFO-SCO equalization for OFDM-RoF in multi-band mobile fronthaul, in: 2020 Opto-Electronics and Communications Conference (OECC), 2020, pp. 1–3. https://doi.org/10.1109/OECC48412.2020.9273585

[38] J. Zhao, A. Yan, G. Li, Z. Li, W. Shen, Y. Hu, N. Chi, Z. He, J. Zhang, et al., 1.92-Tb/s CPRI-equivalent rate direct detection transmission based on ANN pre-equalization for digital-analog radio-over-fiber mobile fronthaul, in: 2024 Optical Fiber Communications Conference and Exhibition (OFC), 2024, pp. 1–3. https://ieeexplore.ieee.org/document/10527151.

[39] J.J.G. Torres, A.M.C. Soto, N.G. González, et al., A novel dispersion monitoring technique in w-band radio-over-fiber signals using clustering on asynchronous histograms, Ingeniería e Investigación 34 (3) (2014) 76–80. https://doi.org/10.15446/ing.investig.v34n3.42902

[40] S. Liu, Y.M. Alfadhli, S. Shen, M. Xu, H. Tian, G.-K. Chang, A novel ANN equalizer to mitigate nonlinear interference in analog-RoF mobile fronthaul, IEEE Photon. Technol. Lett. 30 (19) (2018) 1675–1678. https://doi.org/10.1109/LPT.2018.2865529

[41] D. Zabala-Blanco, M. Mora, C.A. Azurdia-Meza, A. Dehghan Firoozabadi, Extreme learning machines to combat phase noise in RoF-OFDM schemes, Electronics 8 (9) (2019) 921. https://doi.org/10.3390/electronics8090921

[42] C.-C. Wei, Y.-J. Huang, Z.-X. Xie, P.-Y. Huang, P.-H. Ting, C.-T. Lin, \Textgreater100 gbps 3 × 3 MIMO v-band RoF system for up to 100 m wireless transmission enabled by NN-based equalization, in: 2022 Optical Fiber Communications Conference and Exhibition (OFC), 2022, pp. 1–3. https://ieeexplore.ieee.org/document/9748418.

[43] C. Liu, C. Wang, W. Zhou, F. Wang, M. Kong, J. Yu, 81-GHz W-band 60-Gbps 64-QAM wireless transmission based on a dual-GRU equalizer, Opt. Express 30 (2) (2022) 2364–2377. https://doi.org/10.1364/OE.448845

[44] J. Zhang, W. Zhou, C. Bian, J. Ge, S. Xu, Y. Ma, Q. Wang, S. Wang, Z. Ou, S. Hu, 85-GHz Photonics w-band IM/DD PAM4 wireless transmission over 300 m based on

nonlinear u-net symmetrical encoder-decoder equalizer, Opt. Commun. 577 (2025) 131423. https://doi.org/10.1016/j.optcom.2024.131423
[45] S. Xu, B. Sang, L. Zeng, L. Zhao, Two-lane DNN equalizer using balanced random-oversampling for w-band PS-16QAM RoF delivery over 4.6 km, Sensors 23 (10) (2023) 4618. https://doi.org/10.3390/s23104618
[46] Y. Jiang, S. Xu, Q. Wang, J. Zhang, J. Ge, J. Lin, Y. Ma, S. Wang, Z. Ou, W. Zhou, Demonstration of 50 Gbps long-haul d-band radio-over-fiber system with 2D-convolutional neural network equalizer for joint phase noise and nonlinearity mitigation, Sensors 25 (12) (2025) 3661. https://doi.org/10.3390/s25123661
[47] T. Xie, J. Yu, 4Gbaud PS-16Qam d-band fiber-wireless transmission over 4.6 km by using balance complex-valued NN equalizer with random oversampling, Sensors 23 (7) (2023) 3655. https://doi.org/10.3390/s23073655
[48] W. Zhou, L. Zhao, J. Zhang, K. Wang, J. Yu, Y.-W. Chen, S. Shen, R.-K. Shiu, G.-K. Chang, 135-GHz D-band 60-Gbps PAM-8 wireless transmission employing a joint DNN equalizer with BP and CMMA, J. Lightwave Technol. 38 (14) (2020) 3592–3601. https://doi.org/10.1109/JLT.2020.2979070
[49] C. Wang, K. Wang, Y. Tan, F. Wang, B. Sang, W. Li, W. Zhou, J. Yu, High-speed terahertz band radio-over-fiber system using hybrid time-frequency domain equalization, IEEE Photon. Technol. Lett. 34 (11) (2022) 559–562. https://doi.org/10.1109/LPT.2022.3171776
[50] D. Zabala-Blanco, C.A. Azurdia-Meza, A. Dehghan Firoozabadi, P. Palacios Játiva, M. Flores-Calero, I. Soto, S. Kamal, Phase-noise compensation for QPSK-RoF-OFDM signals with the extreme learning machine algorithm for multilayer perceptron, in: 2021 IEEE Latin-American Conference on Communications (LATINCOM), 2021, pp. 1–6. https://doi.org/10.1109/LATINCOM53176.2021.9647840
[51] L.A.M. Pereira, L.L. Mendes, C.J.A. Bastos-Filho, A.C. Sodré, et al., Novel machine learning linearization scheme for 6G a-RoF systems, J. Lightwave Technol. 41 (23) (2023) 7245–7252. https://doi.org/10.1109/JLT.2023.3304281
[52] M. Alharbi, A. Alhuseini, A. Ragheb, M. Altamimi, T. Alshawi, S. Alshebeili, Automatic modulation classification: investigation for millimeter wave over fiber channels, IEEE Photon. Technol. Lett. 31 (13) (2019) 1092–1095. https://doi.org/10.1109/LPT.2019.2918924
[53] T. Alrubeaan, K. Albagami, A. Ragheb, S. Aldosari, M. Altamimi, S. Alshebeili, An investigation of LPI radar waveforms classification in RoF channels, 2019, Pages: 124844–124853 Publication Title: IEEE Access Volume: 7. https://doi.org/10.1109/ACCESS.2019.2938317
[54] R. Saiyyed, M. Sindhwani, S. Sachdeva, H. Pahuja, M.K. Shukla, Optoelectronic equipment-based fault monitoring with 64QAM-OFDM RoF transmission, J. Opt. (2024) 1–12. https://doi.org/10.1007/s12596-024-02226-w
[55] S. Zhang, F. Yaman, K. Nakamura, T. Inoue, V. Kamalov, L. Jovanovski, V. Vusirikala, E. Mateo, Y. Inada, T. Wang, et al., Field and lab experimental demonstration of nonlinear impairment compensation using neural networks, Nat. Commun. 10 (1) (2019) 3033. Publisher: Nature Publishing Group, https://doi.org/10.1038/s41467-019-10911-9
[56] A. Pryamikov, Deep learning as a highly efficient tool for digital signal processing design, Light Sci. Appl. 13 (1) (2024) 248. Publisher: Nature Publishing Group, https://doi.org/10.1038/s41377-024-01599-8
[57] Q. Fan, C. Lu, A.P.T. Lau, et al., Combined neural network and adaptive DSP training for long-haul optical communications, J. Lightwave Technol. 39 (22) (2021) 7083–7091. https://doi.org/10.1109/JLT.2021.3111437
[58] V. Oliari, S. Goossens, C. Häger, G. Liga, R.M. Bütler, M. van den Hout, S. van der Heide, H.D. Pfister, C. Okonkwo, A. Alvarado, et al., Revisiting efficient multi-step nonlinearity compensation with machine learning: an experimental demonstration, J. Lightwave Technol. 38 (12) (2020) 3114–3124. https://doi.org/10.1109/JLT.2020.2994220
[59] J. Zhao, Y. Liu, T. Xu, J. Zhao, Y. Liu, T. Xu, et al., Advanced DSP for coherent optical fiber communication, Appl. Sci. 9 (19) (2019). Company: Multidisciplinary Digital Publishing Institute Distributor: Multidisciplinary Digital Publishing Institute Institution: Multidisciplinary Digital Publishing Institute Label: Multidisciplinary Digital Publishing Institute Publisher: publisher, https://doi.org/10.3390/app9194192
[60] L. Minelli, Advanced digital signal processing and artificial neural networks in modern optical communications, 2025. https://www.semanticscholar.org/paper/Advanced-Digital-Signal-Processing-and-Artificial-Minelli/84af4c76446eeaabb0875ef74af6fe5871d2059b?utm_source=consensus.
[61] Z. Niu, H. Yang, L. Li, M. Shi, G. Xu, W. Hu, L. Yi, et al., Learnable digital signal processing: a new benchmark of linearity compensation for optical fiber communications, Light Sci. Appl. 13 (1) (2024) 188. Publisher: Nature Publishing Group, https://doi.org/10.1038/s41377-024-01556-5