International Journal of Electrical Engineering and Sustainability (IJEES)
ISSN (online): 2959-9229
https://ijees.org/index.php/ijees/index
Arab Impact Factor: 1.51 Dolj 2026 = Q1
Volume 4 | Number 2 | July-December 2026 | Pages 153-170

Deep Learning for Channel Estimation and CSI Feedback in FDD Massive MIMO: A Critical Review and Future Directions

Fadia Massoud Abusalah[1], Abdulgani Albagul[2,3*]
[1]Department of Electrical and Electronics Engineering, Wadi Al-Shati University, Wadi Al-Shati, Libya
[2]Libyan Authority for Scientific Research, Tripoli, Libya
[3]College of Electronic Technology Baniwalid, Baniwalid, Libya

Email: [1]f.massiud@wau.edu.ly, [2]albagoul@yahoo.com

Received: June 02, 2026 Accepted: August 15, 2026 Published: September 03, 2026
DOI: 10.65998/ijees.v4i2.184

Abstract: Accurate channel state information (CSI) is essential for achieving high spectral efficiency in frequency-division duplex (FDD) massive multiple-input multiple-output (MIMO) systems. However, the increasing dimensionality of massive-MIMO channels creates substantial overhead in downlink channel estimation and uplink CSI feedback. This paper presents a structured critical review of deep-learning-based channel estimation and CSI-feedback methods for FDD massive MIMO. A two-level taxonomy first distinguishes channel estimation from CSI feedback and then categorises the reviewed methods according to their functional objectives and neural architectures. Quantitative comparisons are restricted to studies with sufficiently compatible tasks, channel models or datasets, compression ratios, feedback-bit budgets, and evaluation metrics. Channel extrapolation, beam and blockage prediction, vehicular channel prediction, end-to-end transceiver learning, and over-the-air federated learning are discussed separately as adjacent tasks rather than CSI-feedback methods. The article identifies persistent challenges in benchmark standardisation, finite-bit quantisation, feedback-channel reliability, hardware-aware encoder design, computational complexity, cross-domain generalisation, and integration with beam management. Future research directions include physics-informed learning, online and continual adaptation, channel foundation models, and carefully validated generative approaches for CSI acquisition and feedback.

Keywords: Channel state information, deep learning, massive MIMO, channel estimation, CSI feedback, CSI compression, 5G/6G.

1. Introduction

Massive multiple-input multiple-output (MIMO) systems employ large antenna arrays to improve spectral efficiency, energy efficiency, and spatial multiplexing capability [1–3]. Accurate channel state information (CSI) is essential for exploiting these gains, particularly for downlink beamforming and precoding. In time-division duplex (TDD) systems, channel reciprocity can support downlink CSI acquisition from uplink pilots, subject to reciprocity calibration. In frequency-division duplex (FDD) systems, the user equipment (UE) must estimate the downlink channel and convey the resulting CSI to the base station through a rate-limited uplink feedback link [4]. As the number of antennas, subcarriers, and channel dimensions increases, the associated CSI acquisition and feedback overhead can become substantial [4,5].

Conventional CSI acquisition and feedback methods include codebook-based feedback and compressed-sensing techniques. Compressed sensing exploits sparsity or low-dimensional structure in suitable channel representations and can reduce the number of measurements required for channel recovery [5-7]. Its performance, however, depends on the validity of the assumed channel structure and recovery model. Iterative reconstruction can also introduce computational complexity. These limitations have motivated learning-based approaches that can learn nonlinear mappings directly from channel data.

Deep learning (DL) has therefore emerged as an important approach to CSI estimation, compression, reconstruction, and feedback. CsiNet introduced an autoencoder-based framework for massive-MIMO CSI feedback and demonstrated that a learned encoder-decoder can substantially reduce feedback overhead while maintaining useful reconstruction quality [8]. Subsequent studies extended this approach to multiple compression rates [9], time-varying channels [10], model-driven reconstruction [11], quantised feedback [12,13], and lightweight or deployment-oriented architectures [14,15].

These developments have established a broad family of neural approaches for reducing CSI feedback overhead while addressing practical constraints such as feedback rate, computational complexity, and hardware resources. Deep learning has also been applied to channel-estimation and channel-learning problems that are related to, but distinct from, CSI feedback. Examples include deep-learning-based channel parameter acquisition for terahertz ultra-massive MIMO systems [16], uplink channel estimation [17], cross-band channel extrapolation [18], and V2V channel prediction [19]. Learning-based wireless systems have also incorporated over-the-air federated learning in cell-free massive MIMO networks [20,21]. These studies provide important insights into the broader use of learning in wireless communications. However, their objectives, signal models, datasets, and evaluation metrics differ from those of CSI feedback. Combining their reported results directly with CSI-feedback methods can therefore produce misleading conclusions.

Several articles have already examined deep learning for CSI acquisition and feedback. Mashhadi and Gündüz [22] provided an early review covering DL-based massive-MIMO channel acquisition and feedback, including channel estimation and feedback architectures. Guo et al., [23] subsequently presented a comprehensive review focused on DL-based CSI feedback in massive MIMO systems, covering neural architectures, multirate and bit-level feedback, imperfect feedback, complexity, datasets, online training, generalisation, and standardisation [23]. Broader reviews have also examined the use of deep learning in physical-layer communications [24,25]. These studies provide an important foundation for the field, but the rapid development of learning-based CSI methods has produced a diverse literature in which channel estimation, feedback compression, prediction, extrapolation, and end-to-end learning are often evaluated under substantially different assumptions.

This article therefore adopts a task-oriented and condition-aware approach, with primary emphasis on FDD massive-MIMO CSI estimation and feedback. General MIMO channel-estimation studies and adjacent CSI-processing tasks are included when they provide relevant architectural or methodological context, but they are analysed separately from direct FDD CSI-feedback compression. Channel estimation and CSI feedback are treated as distinct learning problems, while channel prediction, cross-band extrapolation, end-to-end transceiver learning, and over-the-air federated learning are treated as adjacent research areas rather than direct CSI-feedback methods. Quantitative results are compared only when the task definition, channel model, compression ratio, and evaluation metric are sufficiently compatible. This approach improves the traceability and interpretability of reported results and reduces unsupported comparisons across heterogeneous experimental settings. The review also considers quantised feedback, hardware-aware encoder and decoder design, computational complexity, cross-domain generalisation, benchmark standardisation, and integration with beam management.

The main objective of this article is therefore not to identify a universally superior neural architecture, but to establish a task-oriented and condition-aware framework for interpreting reported results. The main contributions of this paper are as follows:

1. A unified MIMO signal model and formulation of the learning objectives associated with channel estimation and CSI reconstruction are presented to provide a common basis for analysing the literature.

2. A two-level taxonomy is developed that first separates channel estimation from CSI feedback and then categorises existing approaches according to their functional objectives and neural architectures.
3. Source-verified comparison tables are provided that preserve each study's task definition, channel model or dataset, compression ratio where applicable, evaluation metric, and reported conditions. Results are not directly compared when these conditions are incompatible.
4. A critical assessment is provided of persistent challenges in quantised feedback, hardware-aware model design, computational complexity, cross-domain generalisation, benchmark standardisation, and beam-management integration. Emerging research directions, including physics-informed learning, online adaptation, channel foundation models, and generative channel estimation and feedback, are also discussed.

1.1 Review Methodology and Literature Search

A structured literature search was conducted to identify, select, classify, and critically synthesise studies on deep-learning-based MIMO channel estimation and CSI feedback following the PRISMA 2020 reporting recommendations [26]. The review primarily focuses on massive MIMO and FDD systems, where accurate channel knowledge and limited-rate CSI feedback are important for downlink transmission. Related learning-based channel tasks were also considered when they provided relevant methodological context, but they were treated separately from direct channel-estimation and CSI-feedback studies.

The literature search was conducted using IEEE Xplore, Scopus, Web of Science, ScienceDirect, and Google Scholar. The search covered studies addressing MIMO channel estimation, massive MIMO, channel state information, CSI feedback, CSI compression, and deep-learning-based methods. Search terms included combinations of "MIMO channel estimation," "massive MIMO," "channel state information," "CSI feedback," "CSI compression," "deep learning," "neural network," "autoencoder," "CNN," "LSTM," "transformer," and "quantization." Foundational studies and relevant methodological works were also retained when necessary to establish the technical background, system models, learning architectures, channel models, or evaluation framework of the article. The searches were conducted in Jun and July 2026. The search covered publications within the period from 2016 to 2026.

The search identified 100 records across the five databases. After removal of 25 duplicate records, 75 records remained for title and abstract screening. Following this screening stage, 64 full-text reports were assessed for eligibility. Of these, 43 reports were excluded because they were outside the scope of the review. The 21 studies included in the final synthesis were selected primarily from peer-reviewed literature, while foundational references, datasets, standards, and methodological works were retained separately for technical context. The two authors participated in the study-selection process and reached agreement on the final set of included studies. The literature identification and selection process is summarised in Figure 1.

The review primarily considered peer-reviewed studies addressing deep-learning methods for MIMO channel estimation and CSI feedback. Foundational works, datasets, channel models, and relevant standards were retained separately when necessary to establish the technical background or evaluation framework. Studies addressing related but distinct tasks, including channel prediction, cross-band extrapolation, beam or blockage prediction, end-to-end transceiver learning, and over-the-air federated learning, were considered as adjacent research areas but were not treated as direct CSI-feedback methods unless their results were explicitly relevant to the review scope.

The selected studies were classified according to the primary learning task, channel model or dataset, neural architecture, feedback or compression setting, and reported evaluation metrics. Numerical results were compared only when the task definition, channel model or dataset, compression ratio where applicable, and evaluation metric were sufficiently compatible. Results obtained under different experimental conditions were retained for qualitative discussion rather than combined into a single numerical ranking. This condition-aware approach preserves the context of each reported result and reduces potentially misleading cross-study comparisons.

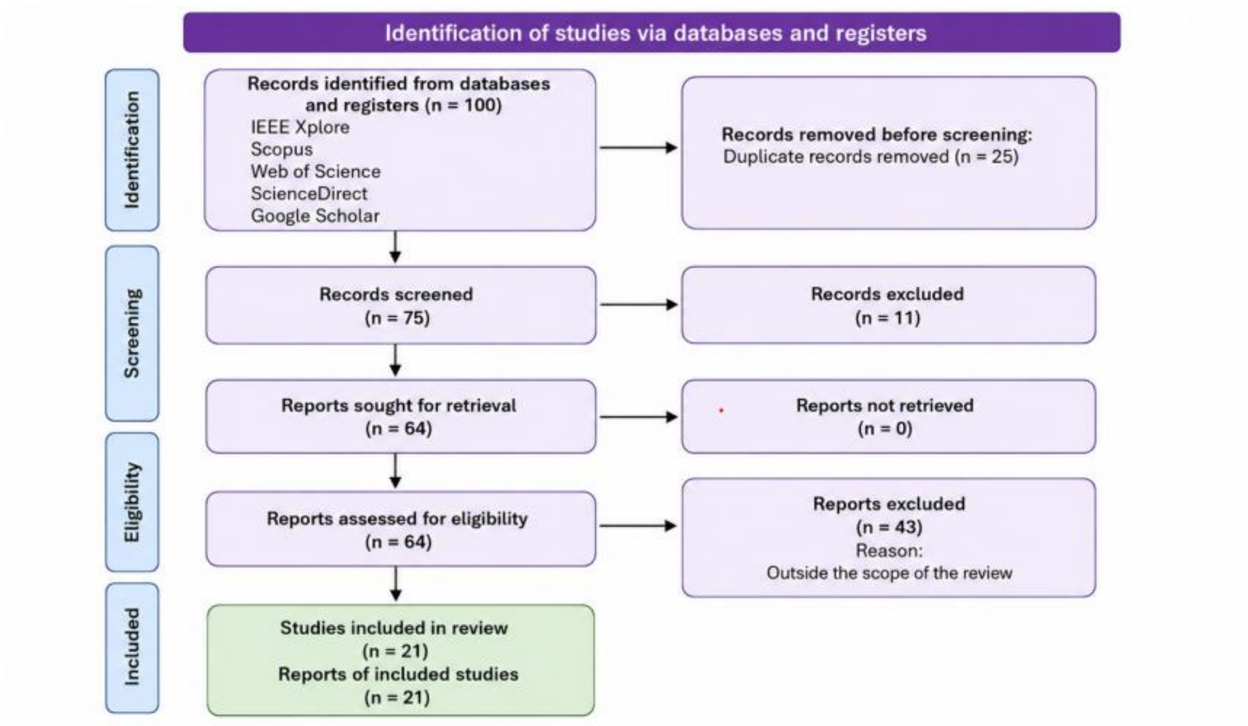

Figure 1. Literature identification and screening flow.

The remainder of this paper is organised as follows. Section II introduces the MIMO signal model and formulates the channel estimation and CSI feedback problems. Section III reviews the principal deep-learning architectures used in these tasks. Sections IV and V present the taxonomies for channel estimation and CSI feedback, respectively. Section VI provides a condition-aware comparative analysis. Section VII discusses open research challenges and future directions. Section VIII concludes the paper.

2. MIMO system model and CSI framework

2.1 Signal model

Consider a narrowband FDD MIMO system with N_t transmit antennas at the base station (BS) and N_r receive antennas at the UE [4]. The received signal vector $y \in \mathbb{C}^{N_r}$y is

$$y = Hx + n \tag{1}$$

where $H \in \mathbb{C}^{N_r N_t}$ is the downlink channel matrix, $x \in \mathbb{C}^{N_t}$ is the transmitted vector, and $n \sim \mathcal{CN}(0, \sigma^2 I)$ is additive white Gaussian noise.

For wideband OFDM with N_c subcarriers, the frequency-domain channel is represented by $\{H[k]\}_{k=0}^{N_c-1}$ or by a tensor $H \in \mathbb{C}^{N_r N_t N_c}$. Applying spatial DFTs gives an angular representation, and a DFT across subcarriers gives the delay dimension [8]. If $N_\tau$delay taps are retained, the angular-delay representation has dimensions $N_r N_t\, N_\tau$, and N_τ must be stated when defining the encoder input dimension.

$$\widetilde{H} = F_r H_0 F_t^H \tag{2}$$

Here, F_r and F_t denote unitary spatial DFT matrices. For OFDM, a DFT across frequency must also be applied before referring to an angular-delay representation; the retained number of delay taps is therefore part of the model definition.

2.2 CSI acquisition and pilot-based estimation

The BS transmits a known pilot matrix $P \in \mathbb{C}^{N_t \tau^P}$, and the UE receives [22 ,23].

$$Y_p = HP + N_p \tag{3}$$

where N_p is pilot noise. When PP^H is nonsingular, the least-squares (LS) estimate is given by (4). An MMSE estimator additionally requires a consistent vectorised-channel covariance model [22]; equation (5) represents the corresponding linear shrinkage form under the stated Gaussian assumptions.

$$\widehat{H}_{LS} = Y_p P^H (PP^H)^{-1} \tag{4}$$

$$\widehat{\mathrm{H}}_{\mathrm{MMSE}} = \mathrm{R}_h \left(\mathrm{R}_h + \frac{\sigma_n^2}{\rho}\mathrm{I}\right)^{-1} \widehat{\mathrm{H}}_{\mathrm{LS}} \tag{5}$$

The covariance $R_h = E\,[hh^H]$ is defined for the vectorised channel h. MMSE optimality holds under the assumed second-order Gaussian model; a learned estimator instead approximates the mapping from pilots or a coarse estimate to CSI using training data and must be validated under distribution shift.

2.3 Distinction among CSI-related learning tasks

To avoid ambiguity in the literature, it is important to distinguish channel estimation, CSI feedback, channel prediction, and channel extrapolation. Although these tasks may employ similar neural architectures, they differ in their inputs, outputs, objectives, and evaluation criteria. Channel estimation aims to recover the channel from pilot observations. CSI feedback compresses an available CSI representation at the UE and reconstructs it at the BS under a feedback-rate constraint. Channel prediction estimates future channel states from previously observed channel information, whereas channel extrapolation estimates the channel in an unobserved frequency band or other domain from available observations. These tasks are therefore treated separately in this review, and their reported numerical results are not combined in a common performance ranking [22 ,23]..

Over-the-air (OTA) federated learning [20,21] is another adjacent research area that should be distinguished from CSI feedback. In OTA federated learning, multiple UEs or base stations collaboratively train a shared machine-learning model by transmitting gradient updates over the wireless channel. While CSI is needed for channel equalisation and precoding in this process, the primary objective is distributed model training rather than CSI compression and reconstruction. The signal model includes the aggregation of multiple UE updates, which is absent in standard CSI feedback. The evaluation metrics (e.g., learning convergence, communication rounds, test accuracy) differ from NMSE or spectral efficiency. Consequently, OTA FL studies are not directly comparable with CSI-feedback methods, and their reported results are excluded from the quantitative comparisons in this review.

2.4 CSI feedback compression

Let $h \in \mathbb{R}^{\mathrm{D}}$ denote the real-valued channel representation supplied to the encoder, where D includes the real and imaginary components and, for OFDM, the retained delay or subcarrier dimension. The UE computes as following.

$$\mathrm{c} = f_\theta(\mathrm{h}), \quad \mathrm{c} \in \mathbb{R}^M, \quad M \ll D \tag{6}$$

After quantization and uplink transmission, the BS reconstructs the channel representation as:

$$\hat{h} = g_\emptyset(C), \; c = f_\theta(h) \tag{7}$$

The compression ratio is defined relative to the actual real-valued encoder input dimension [8]. For an OFDM representation with N_r receive antennas, N_t transmit antennas, and $N_\tau$retained delay taps, the input dimension is $D = 2N_rN_tN_\tau$ unless another preprocessing or antenna convention is explicitly stated.

$$\mathrm{CR} = \frac{M}{D}, \qquad D = 2N_rN_tN_\tau \tag{8}$$

Reconstruction performance is commonly measured by the normalized mean-squared error (NMSE):

$$\mathrm{NMSE} = \mathbb{E}\left[\frac{\|\mathbf{h}-\hat{\mathbf{h}}\|_2^2}{\|\mathbf{h}\|_2^2}\right], \; NMSE_{dB} = 10\,log10(NMSE) \tag{9}$$

The compression ratio CR describes the size of the latent representation relative to the encoder input dimension and is not equivalent to the number of feedback bits. A finite-bit system should report the total feedback budget B separately.

The encoder–decoder pair (f_θ, g_ϕ) is commonly trained end to end by minimising the reconstruction objective:

$$\mathcal{L}_{\theta,\phi} = \mathbb{E}\left[\| \mathrm{h} - g_\phi\big(f_\theta(\mathrm{h})\big) \|_2^2\right] \tag{10}$$

Quantised systems additionally require a discrete representation of the encoder output [12,13]. If the quantiser produces B total feedback bits, the transmitted codeword is represented by q(c) ∈ {0,1}^B.

The reconstructed CSI therefore depends on reconstruction distortion, quantization distortion, and, when present, errors introduced by the uplink feedback channel [12 ,13].

2.5 System architecture overview

Figure 2 illustrates the FDD massive-MIMO processing chain: the BS transmits downlink pilots, the UE estimates the channel, encodes and quantizes the CSI, and sends the finite-bit representation over the uplink feedback channel. The BS demodulates the feedback, decodes the CSI, and uses the reconstructed CSI for precoding. Estimation error, quantization error, and uplink feedback errors are distinct impairments and should be reported separately [8,23].

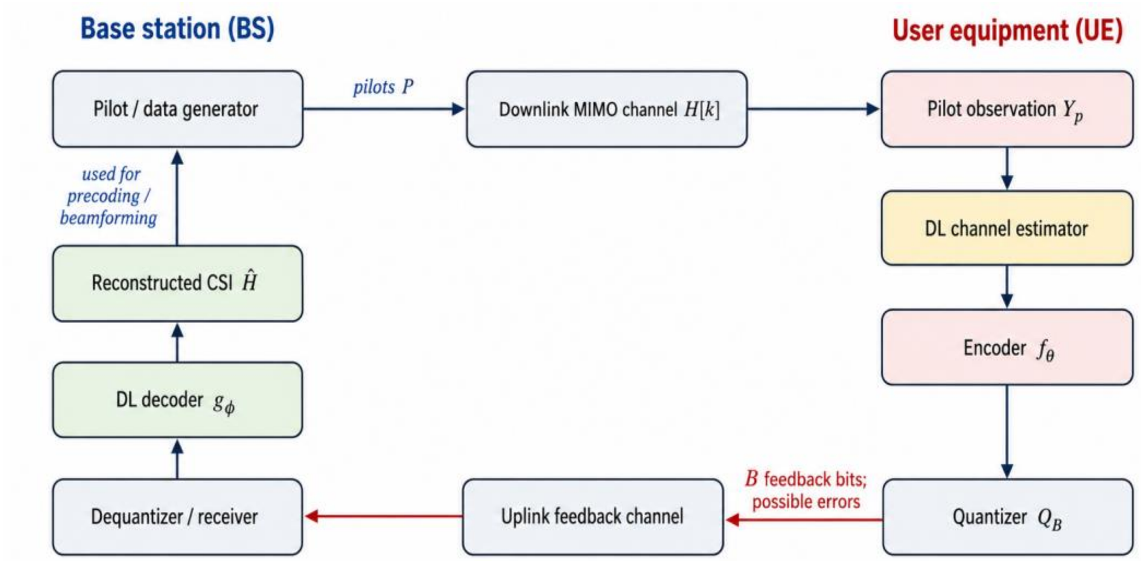

Figure 2. FDD massive-MIMO processing chain for DL-based CSI estimation and feedback.

3. Deep learning architectures for CSI processing

3.1 Convolutional neural networks

CNNs exploit local spatial correlations through shared convolutional kernels and are therefore useful when channel coefficients have structured dependence across antenna, frequency, or delay dimensions [27]. CNN architectures [28] and residual connections [29] provide general architectural foundations, while downsampling, fully connected bottlenecks, and upsampling layers determine the actual compression and computational cost. Architecture descriptions should distinguish parameter count, floating-point operations, model storage, and latency.

3.2 Autoencoder architecture

CsiNet [8] converts the retained angular-delay CSI into two real-valued channels. Its encoder uses a lightweight convolutional feature extractor followed by a fully connected bottleneck. The BS decoder expands the codeword with a fully connected layer and applies two RefineNet blocks. This encoder–decoder asymmetry reflects the tighter compute and energy constraints at the UE.As illustrated in Figure 3, the architecture therefore keeps CSI compression relatively lightweight at the UE while assigning the main reconstruction process to the BS.

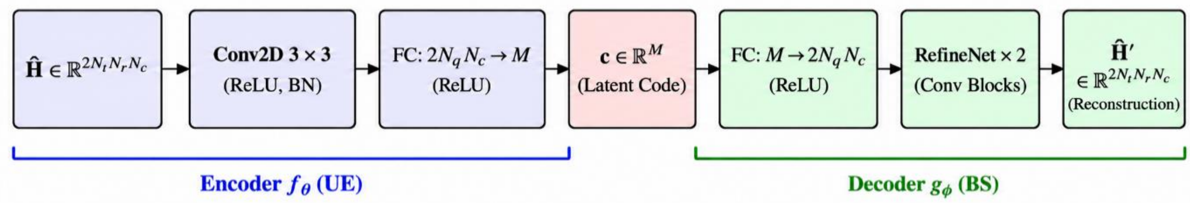

Figure 3. CsiNet autoencoder architecture [8].

3.3 Recurrent neural networks

Recurrent networks, particularly LSTM and GRU, maintain a hidden state and can exploit temporal correlation[30,31]. In CSI processing they are used for sequence feedback, tracking, or prediction, but these tasks are not interchangeable: CsiNet-LSTM [10] is a feedback scheme, whereas the channel tracker in [32] predicts beamspace channels and TC-MIMONet [33] is an end-to-end transceiver evaluated primarily through link performance.

3.4 Transformer-based processing

Multi-head self-attention (MHSA) [34] models long-range dependence among spatial or spectral elements. For query, key, and value matrices Q, K, and V, the attention operation is

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{11}$$

where d_k is the key dimension. Shifted-window attention in the Swin Transformer [35] is one way to limit global-attention cost. Transformer use in wireless systems remains task-dependent: Channelformer [36] addresses channel estimation and online adaptation; Wang et al. [37] present transformer-enabled communication examples including CSI feedback under a bit-constrained setting; and Jia et al. [38] propose the Swin-based channel acquisition network (SCAN) to jointly optimise pilot design, UE-side CSI compression, and BS-side CSI reconstruction. In the same study, a separate SWAN component learns digital-beamforming parameters through a parameterised WMMSE procedure. These components should not be described as one end-to-end joint optimisation of pilots, feedback, reconstruction, and hybrid beamforming. Their results should also not be mapped to a common COST 2100 NMSE baseline unless the experimental frameworks are identical.

3.5 Generative models

Variational autoencoders (VAEs) [39] augment reconstruction with a Kullback–Leiblerregularisation term:

$$\mathcal{L}_{\text{VAE}} = \mathbb{E}\left[\| \text{H} - g_{\phi}(\text{z}) \|_F^2\right] + \beta D_{\text{KL}}\left(q_{\theta}(\text{z} \mid \text{H}) \parallel p(\text{z})\right) \tag{12}$$

where z is the latent variable, $q_{\theta}(z \mid H)$ is an approximate posterior, and β controls the reconstruction–regularisation trade-off. The VAE [39], GAN [40], and learned image-compression references [41] provide general generative and rate–distortion foundations; they do not by themselves establish CSI-feedback gains. Velmurugan et al. [17] apply a diffusion model with a transformer denoiser to uplink MU-MIMO channel estimation, not to CSI feedback compression. Generative claims in feedback therefore require a task-specific source and matched evaluation.

3.6 Deep unfolding and model-driven DL

Deep unfolding converts iterations of an optimisation algorithm into trainable layers. LISTA [42] illustrates the principle, while model-driven CSI feedback in [11] incorporates algorithmic structure into reconstruction. Deep AMP [43] is a sparse-recovery example rather than direct evidence that every unfolded CSI model generalises across channels. Model-driven structure can improve interpretability and constrain the learned reconstruction process, although the resulting robustness and computational benefits remain dependent on the specific architecture and training conditions. Comparisons must state the number of unfolded iterations, trainable parameters, channel distribution, and test mismatch [11, 43,44]. Unfolded networks may still require retraining when array dimensions, pilot structure, or channel statistics change. Their sequential iterations can also increase latency. Configuration-conditioned unfolding and hybrid physics-informed architectures are therefore promising research directions [44].

4. Taxonomy of deep learning-based channel

4.1 Overview and taxonomy

Figure 4 presents the two-level taxonomy used in this review. Level 1 separates channel estimation, CSI feedback, and adjacent CSI-processing tasks. Level 2 groups estimation by observation/design function and feedback by compression strategy. Adjacent tasks are retained for context but excluded from direct feedback-NMSE comparisons. This organization provides a consistent basis for discussing the reviewed methods while avoiding direct comparisons between studies addressing fundamentally different CSI-processing tasks. Channel estimation and CSI feedback are classified separately, while cross-band extrapolation, beam prediction, V2V prediction, and over-the-air federated learning are retained as adjacent CSI-processing tasks. Figure 5 demonstrates selected validated milestones in DL-based CSI processing from 2018 to 2026. These architectures should not be interpreted as a strict performance hierarchy because reported results depend strongly on the channel model, pilot design, SNR, training data, network size, and evaluation protocol. The timeline highlights the progression of CSI estimation and feedback methods from early CNN- and recurrent-based approaches toward model-driven, Transformer-based, quantised, and generative methods, while retaining adjacent CSI-processing studies for context.

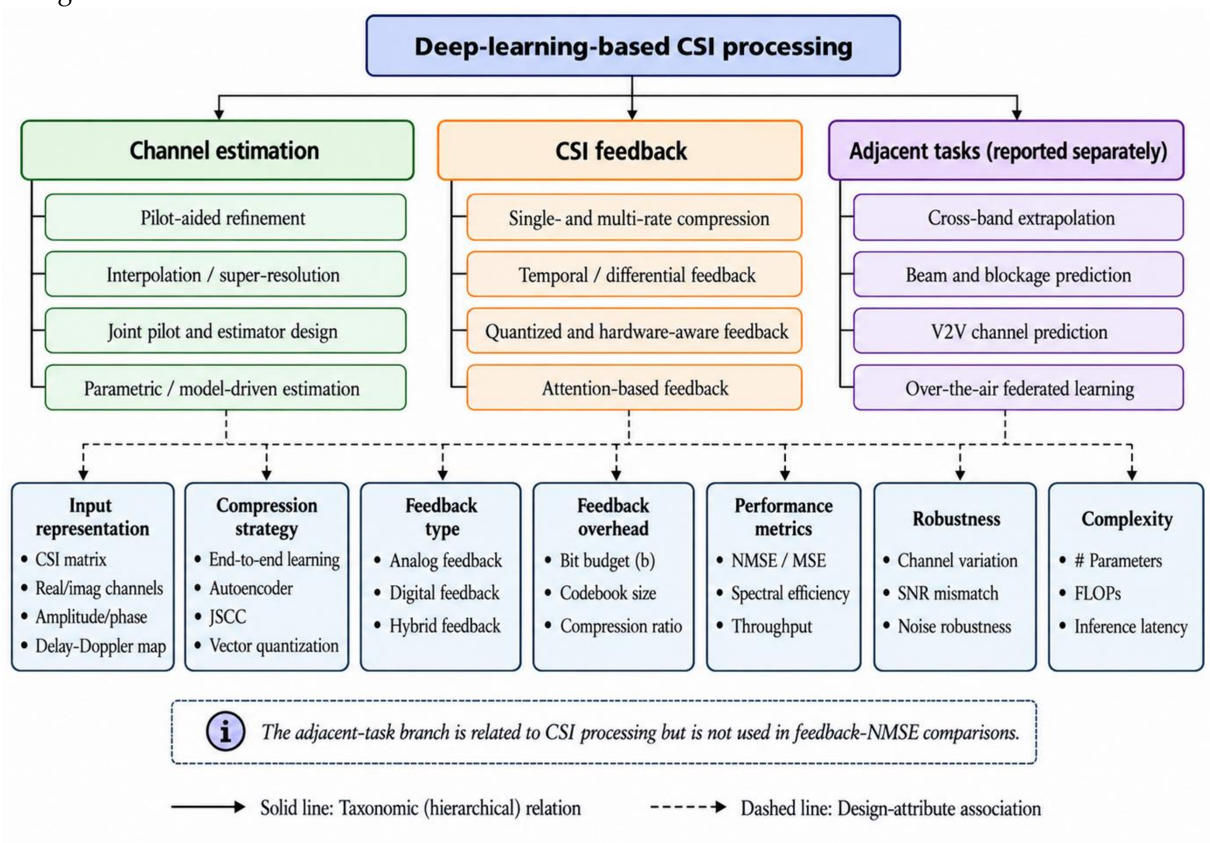

Figure 4. Two-level taxonomy of DL-based CSI processing.

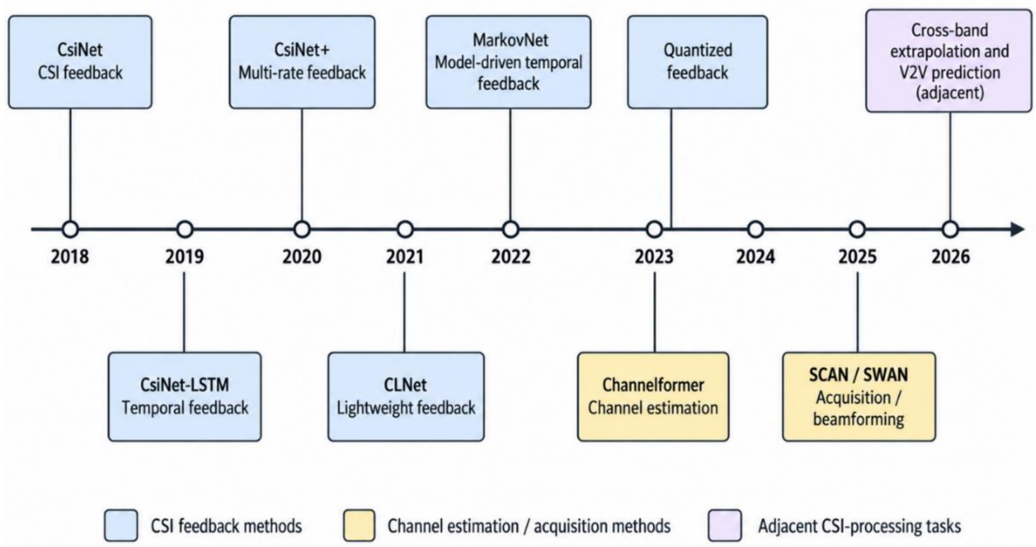

Figure 5. Selected source-verified milestones in deep-learning-based CSI processing (2018–2026). Reference numbers correspond to the bibliography in this manuscript.

4.2 Pilot-aided channel estimation

Pilot-aided estimation maps pilot observations or a coarse estimate to CSI. Dong et al. [45] proposed SF-CNN, SFT-CNN, and SPR-CNN architectures that exploit spatial, frequency, and temporal correlations in mmWave massive MIMO. The method should therefore not be generalized as a universal COST 2100 denoising framework, since its performance is scenario-dependent. The source reports scenario-dependent NMSE and link-level results and shows that SPR-CNN can reduce spatial pilot overhead to approximately one third in its evaluated setting. Mashhadi and Gündüz [46] jointly design frequency-aware pilots and channel estimation using dense pilot-design layers, convolutional processing, non-local attention, and gradual neural pruning. Hu et al. [47] design a low-density frequency-aware antenna-selection pilot pattern together with a transformer-based channel estimator deployed at the UE. Their one-sided design reports performance similar to a two-sided learned pilot/estimator model, saves 50% pilot overhead relative to the conventional method, and remains robust under different antenna-selection patterns. The reported results still depend on the pilot layout, channel model, and SNR, so a single context-free "best NMSE" value is not reported here.

4.3 Interpolation-based estimation

Interpolation-based methods interpret OFDM channel estimation as a super-resolution and denoising problem. Soltani et al. [48] formulate channel estimation as a super-resolution and denoising problem, using a deep-learning architecture to reconstruct the channel from sparse pilot observations. EDSR [49] provides relevant architectural context for the super-resolution component. ChannelNet compares its results with conventional interpolation and MMSE-related baselines under specified OFDM conditions. Channelformer [36] later uses attention and an online-training procedure for wireless channel estimation. Neither wireless paper supplies a universal NMSE that is comparable without matching SNR, pilot pattern, and channel data.

4.4 Joint pilot design and estimation

Joint pilot and estimator design optimises the observation process together with reconstruction. Mashhadi and Gündüz [46] use frequency-aware dense layers for pilot design and gradually prune less significant pilot neurons during training, while convolutional and non-local attention modules exploit channel correlations.

4.5 Transformer-based estimation

Transformer-based estimation uses attention to capture dependencies across pilots, antennas, or subcarriers. Channelformer [36] is an attention-based channel estimator with online training. Kim et al. [16] use a transformer-based parameter-acquisition method for THz ultra-massive MIMO to estimate physical parameters such as angles, distances, and gains from pilots. That study should not be interpreted as a ViT autoencoder for CSI compression and does not support previously reported CR=1/16 feedback results. Transformer architectures have also been explored in combination with model-driven and unfolded frameworks for wireless optimisation problems. While deep unfolding primarily addresses sparse recovery, the combination of attention mechanisms with unfolded structures remains an emerging direction. Further evaluation under consistent pilot density, SNR, channel distributions, and computational budgets is needed to determine when attention-based estimators provide a practical advantage over convolutional or model-driven alternatives.

5. Taxonomy of deep learning-based CSI feedback

5.1 Single-rate convolutional feedback

CsiNet [8] established an influential autoencoder-based framework for learned CSI feedback in massive MIMO and motivated substantial subsequent work on multi-rate compression, temporal feedback, model-driven reconstruction, quantization, and lightweight architectures. On the COST 2100 benchmark at CR=1/4, the paper reports NMSE values of −17.36 dB for the indoor scenario and −8.75 dB for the outdoor scenario. The value −8.65 dB corresponds to the indoor scenario at CR=1/16, not CR=1/4. The paper compares against LASSO, TVAL3, and BM3D-AMP baselines. CsiNet-Plus [50] adds angular-delay truncation and noisy-codeword training. Deep Decoder CsiNet [51] enhances decoder capacity to improve reconstruction performance while maintaining the UE/BS asymmetry. The modification is mainly introduced at the decoder side rather than by increasing encoder complexity [51].

5.2 Multi-rate and adaptive feedback

Guo et al. [9] introduced CsiNet+ together with SM-CsiNet+ and PM-CsiNet+ for multiple-rate feedback and storage reduction. On the COST 2100 benchmark at CR = 1/4, CsiNet+ reports −27.37 dB indoors and −12.40 dB outdoors. CRNet [52] uses multi-resolution feature paths and convolution factorisation rather than dilated convolutions; at the same CR it reports −26.99 dB indoors and −12.71 dB outdoors.

5.3 Temporal and recurrent feedback

CsiNet-LSTM [10] exploits temporal correlation in COST 2100 channel sequences. Its experiments encode the first snapshot at CR=1/4 and evaluate lower ratios for subsequent snapshots; the work is not based on a 3GPP CDL vehicular channel or a universal 200-Hz setting. MarkovNet [53] incorporates a Markovian temporal model into feedback with lower computational burden than recurrent alternatives. The LSTM tracker in [32] and TC-MIMONet [33] are discussed as related tracking/transceiver methods, not as directly comparable feedback compressors.

5.4 Transformer-augmented feedback

Attention can capture long-range channel structure, but transformer results are not uniformly superior under every compression ratio or dataset. Wang et al. [37] evaluate transformer-enabled CSI feedback using a bit-based setup with N_t=64 and K=32 rather than a COST 2100 CR=1/4 benchmark. Jia et al. [38] present SCAN for the joint design of pilots, UE-side CSI compression, and BS-side CSI reconstruction, while a separate SWAN component learns digital beamforming parameters through a parameterised WMMSE procedure. The study therefore should not be represented as a single joint optimisation of pilots, feedback, reconstruction, and hybrid beamforming. Its paper-specific results are not converted into a common NMSE row.

5.5 Lightweight and hardware-efficient

Hardware-efficient feedback must be evaluated using consistent measures. Erak and Abou-Zeid [14] apply pruning, post-training dynamic-range quantization, and weight clustering; they report reductions of 86.5% in model size and 76.2% in inference time while retaining competitive reconstruction. Cui et al. [15] apply the SNIP criterion for single-shot pruning and report comparable or improved reconstruction in their tested setting. CLNet [54] uses a pseudocomplex input representation and spatial attention—not a fully complex network with complex batch normalization—and reports average accuracy improvement of 5.41% with 24.1% lower computational overhead. At CR=1/4, CLNet reports −29.16 dB indoors and −12.88 dB outdoors.

5.6 Differential and incremental feedback

Differential feedback encodes a change relative to a previously reconstructed channel instead of sending a complete representation at every instant. This can exploit temporal correlation but is vulnerable to accumulated reconstruction error. CsiNet-LSTM [10] and the Markovian framework [53] provide temporal/model-driven approaches; the cited sources do not establish a generic learned scheduler that periodically resets the chain, so that claim is omitted.

5.7 Adjacent CSI-related learning tasks

Cross-band extrapolation and other adjacent tasks should be reported separately. MDFCE [18] combines mixture-of-experts and multihead self-attention to map multi-domain sub-6-GHz CSI features to mmWave CSI. Alrabeiah and Alkhateeb [55] address beam and blockage prediction. Chu et al. [19] predict four-dimensional V2V CSI with CA-ConvLSTM, context attention, and meta-learning. The over-the-air federated-learning works [20, 21] optimise distributed learning over cell-free MIMO.

5.8 Quantisation-aware feedback design

Practical feedback requires quantisation of the encoder output. Shao et al. [12] jointly train a compressed-sensing measurement, an ISTA-unfolded decoder, attention, and a vector quantiser. Their numerical results are interpreted only under the reported bit budget and evaluation conditions. The straight-through estimator (STE) [56] is a common heuristic for propagating gradients through discrete operations. Erak and Abou-Zeid [14] study post-training dynamic-range quantisation rather than quantisation-aware training. Vector quantization and learned codebooks are motivated by neural discrete representation learning [57]. Ballé et al. [41] provide a learned image-compression rate–distortion framework, which is a conceptual foundation rather than CSI-specific performance evidence.

Zhou et al. [13] directly study transformer-based CSI feedback with learnable non-uniform quantisation. 3GPP TR 38.843 [58] studies AI/ML for the NR air interface, including CSI-related use cases, but leaves model implementation open.

6. Critical comparative analysis

6.1 Evaluation benchmarks

COST 2100 [59] is a geometry-based stochastic MIMO channel model. The commonly used CsiNet benchmark configurations are 5.3-GHz indoor and 300-MHz outdoor scenarios; those frequencies and preprocessing choices originate from the feedback-study setup rather than defining COST 2100 as a fixed dataset. DeepMIMO [60] is a parameterisable ray-tracing-based dataset framework introduced for millimetre-wave and massive-MIMO applications. Scenario, carrier, antenna geometry, bandwidth, and user sampling must be stated when results are reported. 3GPP CDL models are standardised in TR 38.901 [61] for frequencies from 0.5 to 100 GHz. A CDL result is not directly comparable with COST 2100 unless dimensions, SNR, Doppler, delay spread, preprocessing, and train/test splits are aligned. NMSE, BER, spectral efficiency, and model complexity measure different properties and must not be merged into one ranking.

6.2 Comparison criteria and evidence-compatibility rules

To avoid misleading cross-paper rankings, numerical results are included in the comparative tables only when the compared studies share sufficiently compatible task definitions, channel models or datasets, compression ratios where applicable, and evaluation metrics. When one or more of these conditions differ materially, the reported result is retained as qualitative evidence but is not used to establish numerical superiority. In particular, results from channel estimation, CSI feedback, channel prediction, cross-band extrapolation, beam prediction, and end-to-end transceiver learning are not ranked against one another. For CSI-feedback studies, comparisons are further restricted by compression ratio and evaluation setting whenever these factors materially affect the reported NMSE or reconstruction quality.

6.3 Matched NMSE comparison under common conditions

Figure 6 visualises the four selected COST 2100 results that share CR=1/4 and report both indoor and outdoor NMSE. The figure compares reconstruction accuracy only. Where reported by the source papers, deployment indicators such as parameter count, model-size reduction, and inference-time reduction are retained in the discussion. FLOPs, memory footprint, latency, feedback bits, and spectral efficiency are not consistently reported across these studies and are therefore not used to construct a common complexity ranking.

6.4 Comparison of channel estimation methods

Table 1 summarises representative deep-learning-based channel-estimation methods and their reported evaluation settings. The results show a shift from conventional CNN-based approaches toward attention- and transformer-based methods that exploit spatial, frequency, and temporal channel characteristics. However, differences in channel models, pilot settings, and evaluation metrics prevent a direct numerical ranking of the methods.

6.5 Comparison of CSI feedback methods

Table 2 summarises representative CSI-feedback methods and their main reported findings. The comparison shows the progression from convolutional autoencoders toward multi-rate, temporal, lightweight, quantised, and transformer-assisted approaches. For methods evaluated under the same COST 2100 conditions at CR = ¼, Figure 6 provides a direct NMSE comparison, while the remaining methods are treated as context-specific results.

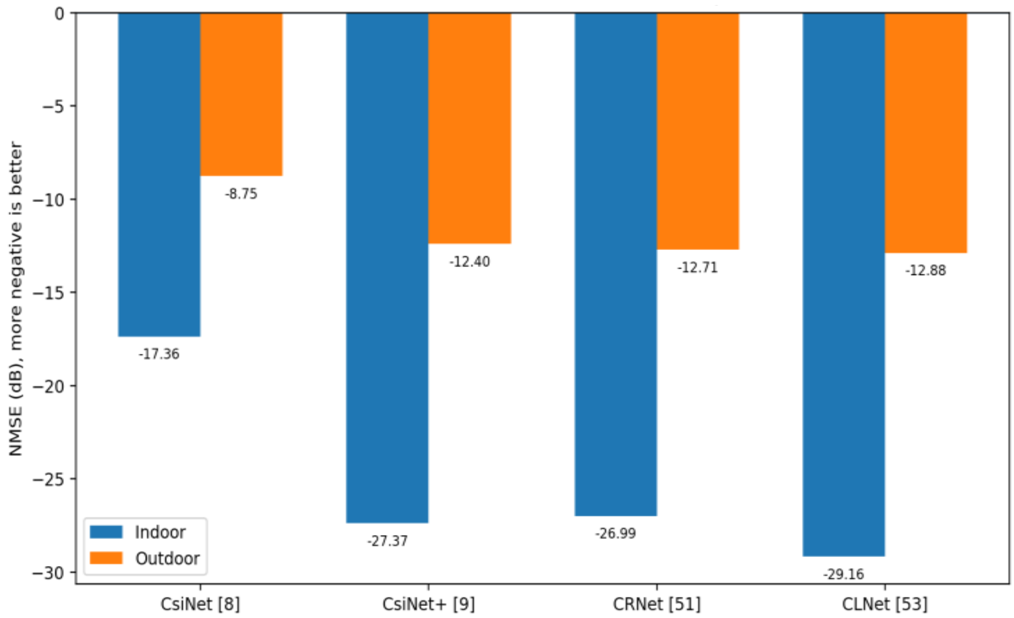

Figure 6. Matched NMSE comparison of selected CSI-feedback methods on COST 2100 at CR=1/4.

Table 1. Comparative Analysis of DL-Based Channel Estimation Methods.

Method	Yr.	Task	Architecture	Scenario	Metric	Source-supported finding	Comparison note
Ye et al. [62]	2018	Joint estimation and detection	Fully connected DNN	OFDM / Rayleigh	BER vs. SNR	Learns the receiver mapping and is evaluated through BER and robustness tests.	Not a feedback-NMSE result.
Soltani et al. [48]	2019	Pilot-based channel estimation / reconstruction	Deep learning-based channel estimation architecture	OFDM multipath channels	NMSE and BER vs. SNR	Formulates channel estimation as a reconstruction problem from sparse pilot observations and evaluates performance under the reported OFDM conditions.	Requires matched pilots, SNR, and channel model.
Dong et al. [45]	2019	mmWave channel estimation	SF-/SFT-/SPR-CNN	mmWave massive MIMO	NMSE, BER, pilot overhead	Exploits spatial, frequency, and temporal correlation; SPR-CNN reduces spatial pilot overhead to about one third in the evaluated setting.	Not COST 2100 feedback.
Mashhadi&Gündüz [46]	2021	Joint pilot design and estimation	Dense pilot design + CNN + non-local attention	FDD MIMO-OFDM	NMSE / pilot overhead	Gradual pruning allocates pilots non-uniformly while exploiting channel correlation.	No context-free peak NMSE.
Channelformer [36]	2023	Channel estimation and online adaptation	Attention-based estimator	Paper-specific OFDM scenarios	NMSE / online adaptation	Uses attention and an online-training procedure for channel estimation.	Not a CSI compressor.
Hu et al. [47]	2023	Low-density pilot channel estimation	Frequency-aware antenna-selection pattern + Transformer estimator	MIMO-OFDM	NMSE vs. pilot/SNR setup	Similar performance to the two-sided learned design; 50% pilot-overhead saving versus the conventional method.	Report antenna-selection pattern, pilot density, channel model, and SNR with every value.
Kim et al. [16]	2023	THz channel-parameter acquisition	Transformer-based parameter estimator	THz ultra-massive MIMO	Parameter/channel estimation error	Estimates physical parameters such as angles, distances, and gains from pilots.	Not CSI-feedback compression.
Velmurugan et al. [17]	2025	Uplink MU-MIMO estimation	DNN regression + diffusion/Transformer	MU-MIMO uplink	MSE vs. SNR	Reports up to 3-dB SNR gain at a target MSE and an additional 0.5 dB at high SNR for the diffusion estimator.	Not CR-based feedback.

Table 2. Comparative Analysis of DL-Based CSI Feedback Methods.

Method	Yr.	Feedback focus	Architecture	Data	Setting	Verified result	Scope / limitation
CsiNet [8]	2018	Single-rate compression	Conv. autoencoder	COST 2100	CR=1/4	Indoor −17.36 dB; outdoor −8.75 dB	Foundational baseline.
CsiNet-LSTM [10]	2019	Temporal feedback	CNN + LSTM	COST 2100 sequences	First snapshot 1/4; later lower CRs	Exploits time correlation; not CDL vehicular.	Sequence setting; not directly comparable to single-snapshot rows.
CsiNet+ family [9]	2020	Multiple-rate feedback	CsiNet+, SM-CsiNet+, PM-CsiNet+	COST 2100	CR=1/4 for matched row	CsiNet+: indoor −27.37 dB; outdoor −12.40 dB	SM/PM variants address storage/rate support.
CsiNet-Plus [50]	2020	Truncation and noisy-codeword training	Convolutional autoencoder	COST 2100	Source-specific	Improves robustness through truncation and noise injection.	No reconstructed universal NMSE.
CRNet [52]	2020	Multi-resolution feedback	Multi-resolution CNN	COST 2100	CR=1/4	Indoor −26.99 dB; outdoor −12.71 dB	Uses multi-resolution paths, not dilated convolution.
MarkovNet [53]	2022	Temporal/model-driven feedback	Markovian model-driven network	Source-specific temporal channels	Source-specific	Exploits channel evolution with lower computation than RNN alternatives.	Do not assign a COST 2100 row without matching setup.
CLNet [54]	2021	Lightweight feedback	Pseudo-complex input + spatial attention	COST 2100	CR=1/4	Indoor −29.16 dB; outdoor −12.88 dB	Reports 24.1% lower computational overhead on average.
Wang et al. [37]	2023	Transformer feedback example	Transformer	Synthetic sparse channel	N_t=64, K=32; bit feedback	Evaluated versus feedback bits and model complexity in the source.	Not COST 2100/CR=1/4.
Deep Decoder CsiNet [51]	2023	Decoder-enhanced feedback	Deeper residual decoder	COST 2100	Source-specific	Improves reconstruction by increasing decoder capacity.	BS-side decoder change; matched value not restated.
Erak Abou-Zeid [14]	2023	Deployment compression	Pruning + post-training quantization + clustering	CsiNet / COST 2100	CR=1/4 experiments	86.5% model-size and 76.2% inference-time reductions.	Not QAT; uses post-training quantization.
SNIP [15]	2023	Single-shot pruning	Pruned feedback network	COST 2100	Source-specific	Reports comparable or improved reconstruction after pruning.	Parameter conventions must be taken from source.
Zhou et al. [13]	2023	Quantised feedback	Transformer + learnable non-uniform quantiser	Massive MIMO	Bit-constrained	Jointly learns CSI feedback and non-uniform quantization.	Use source-specific bit budget.
Jia et al. [38]	2025	CSI acquisition and digital beamforming	SCAN; separate SWAN component	Paper-specific	Two related but distinct components	SCAN jointly optimises pilot design, UE compression, and BS reconstruction; SWAN separately learns WMMSE digital-beamforming parameters.	Do not describe the study as one joint end-to-end optimisation or insert it into the matched NMSE row.

6.6 Critical Analysis of Key Design Choices

Encoder–Decoder Asymmetry: The UE encoder faces tighter memory, energy, and latency constraints than the BS decoder. CsiNet [8] and Deep Decoder CsiNet [51] reflect this asymmetry. Deployment claims should be based on device-level latency, energy, numerical precision, and memory measurements, not parameter count or FLOPs alone. The tables deliberately avoid context-free NMSE and complexity rankings. Results across different channel models, SNRs, pilot patterns, and tasks are not directly ranked. The tables are intended to preserve reported evidence rather than establish a universal ranking of architectures. A higher reported NMSE performance in one row should not be interpreted as superiority unless the corresponding experimental conditions are sufficiently comparable.

For the four matched COST 2100 rows, NMSE values are copied from the cited source tables at CR=1/4. Other rows retain their source-specific evaluation and are explicitly marked non-comparable. Where available, source-reported deployment indicators are retained rather than inferred.

- Fixed vs. Adaptive Compression Ratio: A separate model for every CR increases UE storage. Multiple-rate designs [9] share components or support multiple operating points, but the exact storage and accuracy trade-off depends on which encoders and decoders are shared. Dynamic-rate methods should report switching overhead and the quantised feedback budget.
- Domain Specificity: Many models are trained and tested on one simulated channel distribution. Claims of robustness therefore require cross-scenario tests, such as indoor to outdoor transitions, model to ray tracing shifts, or frequency/array mismatch. Domain adaptation, meta-learning, and online calibration are promising, but they should be evaluated with explicit source/target splits.
- Quantization Compatibility: Infinite-precision bottlenecks overstate deployability. Quantised feedback studies should report the total number of feedback bits, codebook or scalar-quantiser design, training procedure, error-control assumptions, and encoder-device cost [12,14] Post-training compression [14] and quantisation-aware training must be distinguished.
- Scalability to THz and Reconfigurable Intelligent Surfaces Channels above 100 GHz introduce high dimensionality, frequency selectivity, near-field effects, blockage, and hardware constraints [63]. Kim et al. [16] address THz channel-parameter acquisition with transformer-based processing; this does not establish that CNN feedback autoencoders are categorically impractical or that attention complexity is independent of input dimension. Holographic MIMO and reconfigurable intelligent surfaces introduce additional geometry, coupling, and control variables [64]. Future feedback methods should incorporate these physical constraints and evaluate scaling with aperture size, bandwidth, and surface state.

7. Open Research Challenges and Future Directions

Building on the critical assessment in Sections IV–VI, this section identifies key unresolved challenges in the literature and highlights promising research directions. It focuses on areas where further advances are essential to bridge the gap between theoretical developments and practical deployment in 5G-Advanced and 6G systems.

- *Physics-informed and model-driven architectures:* Incorporating physical channel structure (angular-delay sparsity, DFT bases, covariance models) into neural networks could improve sample efficiency and generalisation. Future work should evaluate robustness under channel mismatch and quantify the reduction in training data requirements compared to purely data-driven models.
- *Online and continual adaptation:* Channel statistics change with mobility and varying environments, yet most models are trained on static datasets. Research should explore meta-learning, few-shot adaptation, and unsupervised online calibration, with evaluation under time-varying channel models and explicit reporting of adaptation speed and forgetting.
- *Channel foundation models:* Pre-training on diverse channel datasets and fine-tuning for specific tasks (estimation, feedback, prediction) could reduce task-specific training. Feasibility studies should assess transfer performance across multiple channel models and report zero-shot and few-shot results.
- *Generative approaches:* VAEs, GANs, and diffusion models can represent channel distributions rather than point estimates, potentially improving performance at very low compression ratios. Comparative evaluations against discriminative baselines under identical conditions are needed, with emphasis on downstream task performance (e.g., spectral efficiency) rather than NMSE alone.
- *Quantised end-to-end feedback:* Practical systems require finite-bit feedback. Research should

optimise bit allocation, compare uniform vs. non-uniform and learned vs. fixed quantiser, and quantify the gap between post-training and quantization-aware training.

- *Benchmark standardisation:* The lack of standardised benchmarks hinders fair comparison. We recommend: (i) a minimal reporting standard covering channel model, antenna dimensions, CR, total feedback bits, preprocessing, SNR, and train/test splits; (ii) an open-source benchmark suite; and (iii) code and model release for reproducibility.

Thus, progress toward 5G-Advanced and 6G deployment will depend on addressing these challenges through transparent preprocessing, task-specific metrics, and reproducible hardware measurements.

8. Conclusion

The article separates deep-learning-based channel estimation from CSI feedback and treats cross-band extrapolation, V2V prediction, end-to-end transceiver learning, and over-the-air federated learning as adjacent tasks. The task-oriented taxonomy prevents incompatible NMSE, BER, BLER, and learning-convergence metrics from being combined. The validated COST 2100 results show substantial differences among representative convolutional and attention-assisted designs. The reviewed literature does not support the absolute claim that transformers consistently outperform CNN-based approaches by 2–5 dB across CSI-feedback conditions. Performance depends on channel model, compression ratio, quantization, architecture size, and training/test match. Temporal methods are useful when channel correlation persists, while model compression and quantization are essential for UE deployment.

Three trade-offs remain central: reconstruction accuracy versus feedback and computation; model complexity versus device-level deployability; and specialisation versus cross-domain generalisation. These trade-offs cannot be resolved by NMSE alone. Future work should jointly report feedback bits, encoder latency and energy, memory, robustness to channel mismatch, and downstream spectral efficiency. The matched comparison in Section VI contains only values traceable to the cited source and the same COST 2100/CR=1/4 condition.

Promising directions include physics-informed learning, online adaptation, channel foundation models, and generative approaches, provided their gains are validated against matched baselines. Progress toward 5G-Advanced and 6G deployment will also require standardised data splits, transparent preprocessing, task-specific metrics, and reproducible hardware measurements.

Author Contributions: Authors have contributed significantly to the development and completion of this article.

Funding: This article received no external funding.

Acknowledgments: The authors thank the Wadi Al-Shati University, Libyan Authority for Scientific Research, and College of Electronic Technology Baniwalid, Libya for providing laboratory facilities and equipment.

Ethics approval: The study used public system-level data and involved no human participants.

Conflicts of Interest: The author(s) declare no conflict of interest.

References

[1] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” IEEE Trans. Wireless Commun., vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” IEEE Commun. Mag., vol. 52, no. 2, pp. 186–195, Feb. 2014.

[3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, “An overview of massive MIMO: Benefits and challenges,” IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, pp. 742–758, Oct. 2014.

[4] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, “An overview of limited feedback in wireless communication systems,” IEEE J. Sel. Areas Commun., vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[5] D. L. Donoho, “Compressed sensing,” IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[6] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inf. Theory, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[7] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," Proc. IEEE, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.
[8] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," IEEE Wireless Commun. Lett., vol. 7, no. 5, pp. 748–751, Oct. 2018.
[9] J. Guo, C.-K. Wen, S. Jin, and G. Y. Li, "Convolutional neural network-based multiple-rate compressive sensing for massive MIMO CSI feedback: Design, simulation, and analysis," IEEE Trans. Wireless Commun., vol. 19, no. 4, pp. 2827–2840, Apr. 2020.
[10] T. Wang, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based CSI feedback approach for time-varying massive MIMO channels," IEEE Wireless Commun. Lett., vol. 8, no. 2, pp. 416–419, Apr. 2019.
[11] J. Guo, L. Wang, F. Li, and J. Xue, "CSI feedback with model-driven deep learning of massive MIMO systems," IEEE Commun. Lett., vol. 26, no. 3, pp. 547–551, Mar. 2022.
[12] H. Shao, H. Zhang, W. Zhang, and X. Zhang, "Quantised trainable compressed sensing for MIMO CSI feedback," IEEE Trans. Veh. Technol., vol. 73, no. 12, pp. 19873–19877, Dec. 2024.
[13] B. Zhou, S. Ma, and G. Yang, "Transformer-based CSI feedback with hybrid learnable non-uniform quantization for massive MIMO systems," in Proc. 32nd Wireless Opt. Commun. Conf. (WOCC), 2023, pp. 1–5.
[14] O. Erak and H. Abou-Zeid, "Accelerating and compressing deep neural networks for massive MIMO CSI feedback," in Proc. IEEE Int. Conf. Commun. (ICC), Rome, Italy, 2023, pp. 1029–1035.
[15] Y. Cui, H. Liu, F. Xu, B. Li, and C. Zhao, "Lightweight neural network-based SNIP for CSI feedback in massive MIMO," Electron. Lett., vol. 59, no. 19, Art. No. E12843, Oct. 2023.
[16] S. Kim, A. Lee, H. Ju, K. A. Ngo, J. Moon, and B. Shim, "Transformer-based channel parameter acquisition for terahertz ultra-massive MIMO systems," IEEE Trans. Veh. Technol., vol. 72, no. 11, pp. 15127–15132, Nov. 2023.
[17] P. G. S. Velmurugan, S. S. Ramkumar, S. J. Thiruvengadam, and V. B. Kumaravelu, "Deep learning and diffusion models for uplink channel estimation in MU-MIMO systems," Results Eng., vol. 28, Art. No. 107930, Dec. 2025.
[18] Q. Xiao, K. Li, B. Zhou and S. Ma, "Deep Learning-Based Channel Extrapolation for Dual-Band Massive MIMO Systems," in *IEEE Wireless Communications Letters*, vol. 15, pp. 2994-2998, 2026, doi: 10.1109/LWC.2026.3689267.
[19] L. Chu, D. Burghal, R. Wang, M. E. Neuman and A. F. Molisch, "Context-Conditioned Spatio-Temporal Predictive Learning for Reliable V2V Channel Prediction," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 27, no. 7, pp. 7608-7622, July 2026, doi: 10.1109/TITS.2026.3688571.
[20] H. Sifaou and G. Y. Li, "Over-the-air federated learning over scalable cell-free massive MIMO," IEEE Trans. Wireless Commun., vol. 23, no. 5, pp. 4214–4227, May 2024.
[21] C. Chen, E. Björnson, and C. Fischione, "Multi-task over-the-air federated learning in cell-free massive MIMO systems," arXiv preprint arXiv:2501.17874, 2025.
[22] B. Mashhadi and D. Gündüz, "Deep Learning for Massive MIMO Channel State Acquisition and Feedback," J. Indian Inst. Sci., vol. 100, pp. 369–382, 2020, doi: 10.1007/s41745-020-00169-2.
[23] J. Guo, C.-K. Wen, S. Jin, and G. Y. Li, "Overview of Deep Learning-Based CSI Feedback in Massive MIMO Systems," IEEE Trans. Commun., vol. 70, no. 12, pp. 8017–8045, Dec. 2022, doi: 10.1109/TCOMM.2022.3217777.
[24] T. J. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," IEEE Trans. Cogn. Commun. Netw., vol. 3, no. 4, pp. 563–575, Dec. 2017.
[25] Z. Qin, H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep learning in physical layer communications," IEEE Wireless Commun., vol. 26, no. 2, pp. 93–99, Apr. 2019.

[26] M. J. Page, J. E. McKenzie, P. M. Bossuyt, et al., “The PRISMA 2020 statement: An updated guideline for reporting systematic reviews,” *BMJ*, vol. 372, Art. no. n71, 2021, doi: 10.1136/bmj.n71.

[27] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, May 2015.

[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2012, pp. 1097–1105.

[29] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778.

[30] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[31] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder–decoder for statistical machine translation,” in Proc. EMNLP, 2014, pp. 1724–1734.

[32] H. Liu, Z. Tong, Q. Deng, Y. Zhu, T. Zhang, R. Huang, and Z. Hu, “LSTM-based channel tracking of mmWave massive MIMO systems for mobile Internet of Things,” in Communications, Signal Processing, and Systems (CSPS 2020), Lecture Notes in Electrical Engineering, vol. 654, pp. 529–537, Springer, 2021, doi: 10.1007/978-981-15-8411-4_72.

[33] C. Chen, Z. Wang, Y. Mao, H. Wu, B. Bai, and G. Zhang, “TC-MIMONet: A learning-based transceiver for MIMO systems with temporal correlations,” in Proc. IEEE 93rd Veh. Technol. Conf. (VTC-Spring), 2021, pp. 1–5, doi: 10.1109/VTC2021-Spring51267.2021.9448981.

[34] A. Vaswani et al., “Attention is all you need,” in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2017, pp. 5998–6008.

[35] Z. Liu et al., “Swin transformer: Hierarchical vision transformer using shifted windows,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2021, pp. 10012–10022.

[36] D. Luan and J. S. Thompson, “Channelformer: Attention-based neural solution for wireless channel estimation and effective online training,” IEEE Trans. Wireless Commun., vol. 22, no. 10, pp. 6562–6577, Oct. 2023.

[37] Y. Wang, Z. Gao, D. Zheng, S. Chen, D. Gündüz, and H. V. Poor, “Transformer-empowered 6G intelligent networks: From massive MIMO processing to semantic communication,” IEEE Wireless Commun., vol. 30, no. 6, pp. 127–135, Dec. 2023.

[38] N. Jia, B. Lin, M. Wu, Z. Han, D. Mi, and Z. Gao, “Swin transformer-based massive MIMO channel feedback and hybrid beamforming,” in Proc. 17th IEEE Int. Conf. Wireless Commun. Signal Process. (WCSP), 2025, doi: 10.1109/WCSP68525.2025.1010583.

[39] D. P. Kingma and M. Welling, “Auto-encoding variational Bayes,” in Proc. Int. Conf. Learn. Represent. (ICLR), 2014.

[40] I. J. Goodfellow et al., “Generative adversarial nets,” in Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2014, pp. 2672–2680.

[41] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, “Variational image compression with a scale hyperprior,” in Proc. ICLR, 2018.

[42] K. Gregor and Y. LeCun, “Learning fast approximations of sparse coding,” in Proc. ICML, 2010, pp. 399–406.

[43] Y. Yang, J. Sun, H. Li, and Z. Xu, “Deep AMP for sparse signal recovery,” IEEE Trans. Signal Process., vol. 68, pp. 4733–4747, 2020.

[44] N. Shlezinger, Y. C. Eldar and S. P. Boyd, "Model-Based Deep Learning: On the Intersection of Deep Learning and Optimization," in *IEEE Access*, vol. 10, pp. 115384-115398, 2022, doi: 10.1109/ACCESS.2022.3218802.

[45] P. Dong, H. Zhang, G. Y. Li, I. S. Gaspar, and N. NaderiAlizadeh, “Deep CNN-based channel estimation for mmWave massive MIMO systems,” IEEE J. Sel. Topics Signal Process., vol. 13, no. 5, pp. 989–1000, Sep. 2019.

[46] M. B. Mashhadi and D. Gündüz, “Pruning the pilots: Deep learning-based pilot design and channel estimation for MIMO-OFDM systems,” IEEE Trans. Wireless Commun., vol. 20, no. 10, pp. 6315–6328, Oct. 2021.

[47] R. Hu, C. Hao, Y. Zhang, T. Yoo, J. Namgoong, and H. Xu, “Deep learning-based channel estimation with low-density pilot in MIMO-OFDM systems,” in Proc. IEEE Int. Conf. Commun. (ICC), 2023, pp. 2619–2624, doi: 10.1109/ICC45041.2023.10278650.
[48] M. Soltani, V. Pourahmadi, A. Mirzaei, and H. Sheikhzadeh, “Deep learning-based channel estimation,” IEEE Commun. Lett., vol. 23, no. 4, pp. 652–655, Apr. 2019.
[49] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, “Enhanced deep residual networks for single image super-resolution,” in Proc. IEEE CVPR Workshops, 2017, pp. 136–144.
[50] F. Liu, X. He, C. Li, and Y. Xu, “CsiNet-Plus Model with Truncation and Noise on CSI Feedback,” IEICE Trans. Fundamentals, vol. E103-A, no. 1, pp. 376–381, Jan. 2020, doi: 10.1587/transfun.2019EAL2123.
[51] A. Chakma, S. S. Alam, M. H. Rahman, and Y. M. Jang, “Deep decoder CsiNet for FDD massive MIMO system,” IEEE Wireless Commun. Lett., vol. 12, no. 12, pp. 2073–2077, Dec. 2023.
[52] Z. Lu, J. Wang, and J. Song, “Multi-resolution CSI feedback with deep learning in massive MIMO system,” in Proc. IEEE Int. Conf. Commun. (ICC), 2020, pp. 1–6.
[53] Z. Liu, M. Del Rosario, and Z. Ding, “A Markovian model-driven deep learning framework for massive MIMO CSI feedback,” IEEE Trans. Wireless Commun., vol. 21, no. 2, pp. 1214–1228, Feb. 2022.
[54] S. Ji and M. Li, “CLNet: Complex input lightweight neural network designed for massive MIMO CSI feedback,” IEEE Wireless Commun. Lett., vol. 10, no. 10, pp. 2318–2322, Oct. 2021.
[55] M. Alrabeiah and A. Alkhateeb, “Deep learning for mmWave beam and blockage prediction using sub-6 GHz channels,” IEEE Trans. Commun., vol. 68, no. 9, pp. 5504–5518, Sep. 2020.
[56] Y. Bengio, N. Léonard, and A. Courville, “Estimating or propagating gradients through stochastic neurons for conditional computation,” arXiv:1308.3432, 2013.
[57] A. van den Oord, O. Vinyals, and K. Kavukcuoglu, “Neural discrete representation learning,” in Proc. NeurIPS, 2017, pp. 6306–6315.
[58] 3GPP, “Study on artificial intelligence (AI)/machine learning (ML) for NR air interface,” 3GPP TR 38.843 V18.0.0, Jan. 2024.
[59] L. Liu et al., “The COST 2100 MIMO channel model,” IEEE Wireless Commun., vol. 19, no. 6, pp. 92–99, Dec. 2012.
[60] A. Alkhateeb, “DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications,” arXiv:1902.06435, 2019.
[61] 3GPP, “Study on channel model for frequencies from 0.5 to 100 GHz,” 3GPP TR 38.901 V18.1.0, Jan. 2026.
[62] H. Ye, G. Y. Li, and B.-H. Juang, “Power of deep learning for channel estimation and signal detection in OFDM systems,” IEEE Wireless Commun. Lett., vol. 7, no. 1, pp. 114–117, Feb. 2018.
[63] T. S. Rappaport et al., “Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond,” IEEE Access, vol. 7, pp. 78729–78757, 2019.
[64] C. Huang et al., “Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends,” IEEE Wireless Commun., vol. 27, no. 5, pp. 118–125, Oct. 2020.