International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

Attention-Based Deep Learning for Pilot-Assisted Channel Estimation and Signal Detection in 6G Massive MIMO Networks

M.Hema preethi
ME Communication System
Dept. of Electronics and Communication Engineering.
PET Engineering College,Valliyur, Tirunelveli

Ms.K.Shena,M.E
Assistant professor
Dept. of Electronics and Communication Engineering.
PET Engineering College,Valliyur, Tirunelveli

Abstract: Accurate channel estimation and reliable symbol detection are essential for exploiting the spatial multiplexing gains of massive multiple-input multiple-output (MIMO) wireless systems. This paper investigates an attention-based deep-learning approach for refining pilot-assisted least-squares (LS) channel estimates in an 8 × 8 MIMO simulation. The simulated link employs a Rayleigh fading channel, additive white Gaussian noise (AWGN), and 16-quadrature amplitude modulation (16-QAM). The proposed approach refines the LS estimate using an attention-based neural network, after which zero-forcing (ZF) detection is applied to recover the transmitted symbols. Its performance is evaluated against LS, minimum mean-square-error (MMSE), and feed-forward neural network (FFNN) baselines over signal-to-noise ratios (SNRs) from 0 to 20 dB. The reported experiment uses 3,000 training samples, 500 testing samples, and 15 training epochs. At 20 dB SNR, the attention-based model achieves a bit error rate (BER) of 0.22525 and a symbol error rate (SER) of 0.39663, compared with 0.43219 and 0.68150, respectively, for the FFNN baseline. However, the attention-based model does not outperform MMSE across the evaluated SNR range, and its BER improvement becomes limited at higher SNRs. It also performs worse than LS at the higher evaluated SNR points. These findings indicate that attention-based refinement can improve upon the evaluated FFNN baseline, but the current implementation does not establish superiority over conventional estimators. Further validation using repeated trials, a fully documented network architecture, and controlled comparisons is required to assess its robustness and generalisation.

Keywords: Massive MIMO, 6G wireless communications, channel estimation, signal detection, attention-based deep learning, pilot-assisted transmission, bit error rate, 16-QAM.

I. INTRODUCTION

A. Background

The continued evolution of wireless communication systems is driven by the demand for higher data rates, improved spectral efficiency, reliable connectivity, and support for an increasing number of connected devices. Sixth-generation (6G) networks are expected to extend the capabilities of current wireless technologies to support emerging applications such as immersive communication, intelligent transportation, industrial automation, and large-scale machine-type communication. Achieving these objectives requires improvements not only in radio access technologies but also in the physical-layer algorithms responsible for transmitting and recovering information over wireless channels. Multiple-input multiple-output (MIMO) technology is an important component of this development. By employing multiple antennas at the transmitter and receiver, a MIMO system can transmit several data streams over the same time-frequency resources. Under suitable propagation and channel conditions, spatial multiplexing can increase spectral efficiency without requiring a proportional increase in bandwidth. Massive MIMO extends this principle by using a larger number of antenna elements, enabling additional spatial degrees of freedom and improved control over

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

transmitted and received signals. However, the benefits of MIMO depend strongly on the receiver's knowledge of the propagation channel. Wireless signals experience attenuation, phase rotation, fading, and interference as they propagate between the transmitter and receiver. Channel estimation attempts to characterise these effects so that the receiver can compensate for channel distortion and recover the transmitted information. Errors in the channel estimate can propagate into the signal detection stage, reducing detection reliability and increasing the bit error rate (BER) and symbol error rate (SER). Pilot-assisted channel estimation is a widely used approach in which known reference symbols are transmitted alongside information-bearing symbols. Because the receiver knows the transmitted pilot values, it can use the corresponding received observations to estimate the channel coefficients. The quality of this estimate depends on factors including the signal-to-noise ratio (SNR), pilot configuration, channel characteristics, and estimation algorithm. As the number of antennas increases, the dimensionality of the channel matrix also increases, making accurate and computationally efficient estimation an important design consideration. Conventional estimation and detection techniques provide established baselines for evaluating new approaches. Least-squares (LS) estimation offers a relatively straightforward method for obtaining channel estimates from pilot observations. Minimum mean-square-error (MMSE) estimation incorporates statistical information to reduce estimation error under its modelling assumptions, although its practical implementation can require additional information and computational operations. These methods provide useful reference points for assessing whether a learning-based technique offers a measurable advantage under a specified simulation setting.

B. Motivation

Deep learning has attracted research interest in wireless physical-layer processing because neural networks can learn nonlinear relationships from training data. Depending on their architecture, training procedure, and input representation, learning-based methods may approximate aspects of channel estimation or signal detection that are difficult to express through simple fixed mappings. Such methods are particularly relevant when conventional algorithms face challenging channel conditions or when system designers wish to investigate data-driven alternatives to individual receiver-processing blocks. Feed-forward neural networks (FFNNs) have been investigated for learning-based channel estimation and signal detection. Nevertheless, their performance depends on the information supplied to the network, the network's capacity, the quality and diversity of the training data, and the optimisation procedure. A network that does not learn a useful representation of the channel may fail to exploit changes in the received signal as the SNR increases. Consequently, replacing a conventional estimator with a neural network does not automatically guarantee improved communication performance. Attention mechanisms provide another approach to learning relationships within input representations. An attention-based model can assign different weights to elements of an input representation when constructing an output representation. In wireless signal processing, this capability motivates investigation into whether the model can learn useful relationships among the components of a noisy channel estimate. Attention-based channel-estimation approaches have been explored in prior research [7], providing a basis for investigating their application to the system considered in this study. The present work investigates an attention-based neural network that refines an LS channel estimate obtained from pilot observations. The refined estimate is then used by a zero-forcing (ZF) detector to recover transmitted symbols. This design allows the attention-based estimator to be examined within a defined MIMO simulation pipeline and compared with conventional and neural-network baselines. The objective is not to assume that attention is inherently superior, but to determine how the implemented model performs under the experimental conditions used in this study.

C. Problem Statement

This study considers an 8 × 8 MIMO communication system operating over a Rayleigh fading channel with additive white Gaussian noise (AWGN). The transmitter uses 16-quadrature amplitude modulation (16-QAM), and the receiver performs pilot-assisted channel estimation followed by symbol detection. The investigated SNR range is 0–20 dB. The primary problem is to determine whether an attention-based neural network can refine an LS channel estimate in a manner that improves the recovery of transmitted information compared with selected conventional and learning-based baselines. The evaluation considers LS, MMSE, FFNN, and attention-based methods using BER and SER, together with channel-estimation mean-square error (MSE) where applicable. The problem has several related aspects. First, the estimator must process information derived from the pilot-assisted channel estimate. Second, the resulting channel estimate must be suitable for the downstream detection stage. Third, the performance must be evaluated across multiple SNR values rather than at a single operating point. Finally, the observed differences must be interpreted in relation to the chosen baselines and simulation assumptions.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

A central research question is therefore: To what extent does attention-based refinement of an LS channel estimate improve channel-estimation and signal-detection performance in the simulated 8 × 8 MIMO system, relative to LS, MMSE, and FFNN baselines across the evaluated SNR range? The investigation also considers whether any observed improvement remains consistent as the SNR changes. This is important because a method may perform favourably under one noise condition but provide limited benefit under another. The answer must therefore be established through measured results rather than assumed from the use of an attention mechanism.

D. Research Objectives

The study is guided by the following objectives:

1. Develop an attention-based channel-refinement approach: Investigate an attention-based neural network that processes the LS channel estimate within the simulated MIMO receiver.
2. Evaluate signal-detection performance: Examine how the refined channel estimate affects the recovery of transmitted 16-QAM symbols through the implemented ZF detection stage.
3. Compare against established baselines: Evaluate the attention-based approach alongside LS, MMSE, and FFNN methods under the specified simulation conditions.
4. Measure performance across SNR values: Analyse BER and SER over the evaluated range of 0, 5, 10, 15, and 20 dB, together with channel-estimation MSE where the required measurements are available.
5. Identify performance limitations: Determine the operating conditions under which attention-based refinement improves upon the selected baselines and the conditions under which its performance remains inferior.
6. Establish a reproducible experimental basis: Document the network architecture, data-generation procedure, training configuration, evaluation protocol, and software environment sufficiently to support independent verification of the reported results.

These objectives combine performance evaluation with methodological transparency. A valid comparison must distinguish improvement over a particular baseline from general superiority over conventional estimation and detection methods.

E. Contributions of This Study

The contributions of this study are defined by the implemented system and the results reported in the current experiments.

1. Attention-based LS channel refinement: The study investigates an attention-based neural network as a refinement stage applied to an LS channel estimate in an 8 × 8 MIMO simulation.
2. Evaluation within a receiver-processing pipeline: The refined estimate is used with ZF detection, allowing the effect of the estimation approach to be examined through downstream symbol-recovery metrics.
3. Comparative performance assessment: The implemented approach is evaluated against LS, MMSE, and FFNN baselines using BER and SER across five SNR points from 0 to 20 dB. Channel-estimation MSE is also considered in the reported analysis.
4. Identification of performance trade-offs: The reported results indicate that the attention-based method performs better than the FFNN baseline across the evaluated SNR points and better than LS at the lower evaluated SNR points. However, it does not outperform MMSE, and its BER becomes nearly flat at higher SNRs while LS achieves lower BER.
5. Identification of directions for further validation: The analysis highlights the need for a fully specified attention architecture, repeated experimental trials, and additional controlled evaluations before broader claims about robustness, generalisation, or computational efficiency can be supported.

These contributions should be interpreted within the limits of the current experiment. The work evaluates a particular implementation and simulation configuration; it does not establish that attention-based estimation universally outperforms conventional methods or that the simulated system is representative of every 6G deployment scenario.

F. Organisation of the Paper

The remainder of this paper is organised as follows. Section II introduces the technical background and mathematical formulation of the MIMO channel, pilot-assisted channel estimation, modulation, and signal detection methods considered in the study. Section III reviews related research on conventional channel estimation, learning-based wireless receivers, and attention-based estimation techniques, with emphasis on the relationship between existing approaches

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

and the present investigation. Section IV describes the proposed method, simulation configuration, training procedure, baseline implementations, and performance measures. Section V presents the experimental results and discusses BER, SER, and channel-estimation MSE across the evaluated SNR range. Section VI summarises the principal findings, identifies the limitations of the current implementation, and outlines directions for further investigation.

II. BACKGROUND

A. MIMO System Model

Multiple-input multiple-output (MIMO) communication uses multiple transmit and receive antennas to improve spectral efficiency and support the simultaneous transmission of multiple data streams. In a massive MIMO configuration, the increased number of antennas can provide additional spatial degrees of freedom. However, these benefits depend on the receiver's ability to estimate the wireless channel and recover the transmitted symbols accurately.

Consider a narrowband MIMO system with N_t transmit antennas and N_r receive antennas. The received signal is represented by

$$\mathbf{y} = \mathbf{Hx} + \mathbf{n} \qquad (1)$$

where x ∈ ℂ^(N_t × 1) is the transmitted symbol vector,

H ∈ ℂ^(N_r × N_t) is the channel matrix,

n ∈ ℂ^(N_r × 1) represents additive noise,

and y ∈ ℂ^(N_r × 1) is the received signal vector.

Each element of the channel matrix represents the complex propagation coefficient between a transmit antenna and a receive antenna. Consequently, the channel matrix describes the combined effects of signal attenuation and phase variation across the antenna links. In the 8 × 8 configuration considered in this study, both the number of transmit antennas and the number of receive antennas are eight. The channel matrix therefore contains 64 complex coefficients.

The receiver must account for the channel's influence before the transmitted data can be recovered reliably. Channel estimation provides an approximation of the channel matrix, while signal detection uses the received signal and a channel representation to estimate the transmitted symbols. The accuracy of these two operations directly affects the resulting bit error rate (BER) and symbol error rate (SER).

B. Pilot-Assisted Least-Squares Channel Estimation

Pilot-assisted channel estimation uses known reference symbols to obtain information about the propagation channel. The transmitter sends pilot symbols whose values are known to the receiver. By comparing the received pilots with the corresponding transmitted pilots, the receiver can estimate the channel coefficients.

Let X_p ∈ ℂ^(N_t × T_p) denote the pilot matrix,

where T_p represents the number of pilot observations. The corresponding received pilot matrix is

$$\mathbf{Y}_p = \mathbf{HX}_p + \mathbf{N}_p \qquad (2)$$

where N_p represents the noise affecting the pilot observations.The least-squares (LS) estimator obtains a channel estimate by minimising the squared difference between the received pilot observations and the observations predicted by the estimated channel. Its objective is

$$\hat{\mathbf{H}}_LS = \arg\min_G \|\mathbf{Y}_p - \mathbf{GX}_p\|^2_F \qquad (3)$$

where G is a candidate channel matrix and ||·||_F denotes the Frobenius norm.When the pilot matrix has sufficient rank, the LS estimate can be expressed as

$$\hat{\mathbf{H}}_LS = \mathbf{Y}_p\mathbf{X}_p^H(\mathbf{X}_p\mathbf{X}_p^H)^{-1} \qquad (4)$$

where (·)^H denotes the Hermitian transpose. This expression assumes that X_p X_p^H is invertible.The principal advantage of LS estimation is its simplicity. It does not require prior knowledge of the channel covariance matrix or a detailed statistical description of the channel. However, the estimator does not explicitly account for the noise statistics. Noise present in the received pilots therefore affects the estimated channel coefficients, particularly when the signal-to-noise ratio (SNR) is low. In this study, the LS estimate is also important because it provides the initial channel representation for the attention-based refinement approach. The learned model is intended to refine this estimate rather than replace the entire physical-layer processing chain.

C. MMSE and Zero-Forcing Detection

After channel estimation, the receiver must separate the transmitted data streams from the received signal. Linear detection methods provide practical ways to perform this operation without exhaustively evaluating every possible transmitted symbol vector.

1) Minimum Mean-Square-Error Detection

The minimum mean-square-error (MMSE) detector balances interference suppression against noise

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

enhancement. Under the conventional linear model, assuming a channel matrix H, spatially white noise with variance σ_n², and transmitted symbols with average energy E_s, an MMSE estimate can be written as

$$\hat{\mathbf{x}}_\text{MMSE} = (\mathbf{H}^\text{H}\mathbf{H} + (\sigma_\text{n}{}^2/\text{E}_\text{s})\mathbf{I})^{-1}\mathbf{H}^\text{H}\mathbf{y} \qquad (5)$$

where I is the identity matrix of the required dimension.The regularisation term involving the noise variance helps prevent the detector from excessively amplifying noise when the channel matrix is poorly conditioned. This can make MMSE detection more robust than zero-forcing detection in noisy conditions. The performance of an MMSE detector depends on the assumptions used to construct it, including the channel representation and noise variance. For a fair comparison, the implementation must specify whether the detector uses the true simulated channel matrix or an estimated channel matrix. That detail should be documented in the methodology section according to the actual simulation code.

2) Zero-Forcing Detection

Zero-forcing (ZF) detection attempts to suppress inter-stream interference by applying the pseudoinverse of the channel matrix. For a full-column-rank channel matrix, its linear estimate is

$$\hat{\mathbf{x}}_\text{ZF} = (\mathbf{H}^\text{H}\mathbf{H})^{-1}\mathbf{H}^\text{H}\mathbf{y} \qquad (6)$$

Equivalently, the expression can be represented using the Moore–Penrose pseudoinverse as

$$\hat{\mathbf{x}}_\text{ZF} = \mathbf{H}\dagger\mathbf{y} \qquad (7)$$

where H† denotes the pseudoinverse of the channel matrix.

When the channel is not well conditioned, ZF may amplify noise while suppressing interference. Its performance is also sensitive to channel-estimation errors. Therefore, a ZF detector operating with an imperfect channel estimate can exhibit a substantial error rate even when the receiver noise becomes relatively small. In the proposed simulation, the attention-based model refines the LS channel estimate, and ZF detection is applied to the refined estimate. The quality of the refined channel representation is consequently important to the final detection performance. The exact channel input used by each baseline must be stated consistently in Section IV to ensure that the comparisons are interpretable.

D. 16-QAM Modulation

Quadrature amplitude modulation (QAM) conveys information by varying the amplitude and phase of a complex-valued carrier representation. In 16-QAM, the modulation constellation contains 16 distinct symbol points. Each symbol represents four information bits because

$$\log_2(16) = 4 \qquad (8)$$

At the transmitter, the input bit stream is grouped into four-bit sequences and mapped to constellation symbols. These symbols form the transmitted vector x. After propagation through the MIMO channel and the addition of noise, the receiver estimates the transmitted symbols using its channel representation and detection algorithm. The estimated symbols are subsequently demodulated into bit sequences. Differences between the original transmitted bits and the recovered bits contribute to the BER, while differences between the original and detected symbols contribute to the SER. The use of 16-QAM allows the study to evaluate the combined effects of channel estimation and signal detection under a modulation scheme that carries multiple bits per symbol. Nevertheless, the findings from this configuration should not automatically be generalised to other modulation orders without further experiments.

E. Attention Mechanism

Attention is a neural-network mechanism that learns how to assign different weights to elements of an input representation. Rather than processing all input components with an identical weighting, an attention operation produces a weighted representation based on relationships learned from the data.

A common formulation is scaled dot-product attention:

$$\text{Att}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}(\mathbf{Q}\mathbf{K}^\text{T}/\sqrt{\text{d}_\text{k}})\mathbf{V} \qquad (9)$$

where Q, K, and V denote the query, key, and value representations, respectively, and d_k is the dimensionality of the key representation. The softmax operation converts the compatibility scores into normalised weights.For channel estimation, attention may help a model learn relationships among features derived from noisy channel observations. When used to refine an LS estimate, the learned operation can potentially exploit patterns that a simple feed-forward mapping does not capture as effectively. However, the presence of an attention operation does not, by itself, guarantee improved channel estimation or lower BER. The outcome depends on the network architecture, input representation, training data, optimisation procedure, and test conditions. The actual type of attention, number of attention heads, layer dimensions, and tensor shapes used in this study must therefore be specified from the implementation rather than inferred from the general attention equation.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

F. Feed-Forward Neural Network Baseline

A feed-forward neural network (FFNN) processes its input through a sequence of layers, with information moving from the input layer towards the output layer. Each fully connected layer applies a learned affine transformation followed by an activation function.

A layer can be represented as

$$f(\mathbf{z}^{(l-1)}) = f(\mathbf{W}^{(l)}\mathbf{z}^{(l-1)} + \mathbf{b}^{(l)}) \qquad (10)$$

where z^(l-1) is the input to layer l, W^(l) is its weight matrix, b^(l) is its bias vector, and f(·) is the activation function. The FFNN baseline provides a useful comparison because it learns a nonlinear mapping without explicitly using an attention operation. The base study discussed in the manuscript uses feed-forward networks for channel estimation and signal detection. In the present work, the FFNN serves as a reference for examining whether the attention-based approach improves the evaluated error metrics under the reported simulation conditions. A valid comparison requires the baselines to use clearly defined inputs, outputs, data partitions, and evaluation procedures. Differences in training data, optimisation, or access to channel information could otherwise influence the results independently of the network architecture.

G. Channel and Noise Model

The simulation considers a Rayleigh fading channel with additive white Gaussian noise (AWGN). Under the standard complex Rayleigh-fading model, each channel coefficient is represented by a zero-mean complex Gaussian random variable. The corresponding coefficient magnitude follows a Rayleigh distribution.

For a channel coefficient h_(r,t), a common normalised representation is

$$h_{r,t} \sim \mathcal{CN}(0, 1) \qquad (11)$$

where CN(0, 1) denotes a circularly symmetric complex Gaussian distribution with unit average power.This notation describes the standard normalised model; the variance and normalisation used in the actual code must be verified before it is reported as the implemented configuration. The AWGN model represents random noise added to the received signal. In the complex baseband representation, the noise is commonly modelled as a zero-mean complex Gaussian vector. Its variance determines the noise power and, together with the signal power, defines the SNR.

The SNR in decibels is expressed as

$$\text{SNR_dB} = 10\log_{10}(P_s/P_n) \qquad (12)$$

where (P_s) is the signal power and (P_n) is the noise power under the adopted definition.

This work reports evaluation over an SNR range from 0 dB to 20 dB. At 0 dB, the signal and noise powers are equal under the stated power-ratio definition. At 20 dB, the signal power is 100 times the noise power. Evaluating multiple SNR levels makes it possible to observe whether the relative performance of the estimation and detection methods changes as the noise conditions improve. The precise channel-generation procedure, fading normalisation, noise variance calculation, and SNR step size should be documented in Section IV so that the experiment can be reproduced.

H. Operation Without Prior Channel State Information

Channel state information (CSI) describes the channel characteristics required by a receiver to process the received signal. In practical pilot-assisted systems, the receiver can estimate CSI by using known pilot symbols transmitted through the channel. In this paper, the term without prior CSI should be understood as the absence of a pre-supplied channel estimate for the learned processing stage, rather than the absence of channel information throughout the receiver. The pilot observations provide information from which an LS estimate can be computed, and the attention-based model is intended to refine that estimate. For an (N_t × N_r) antenna configuration, the channel matrix contains (N_r × N_t) complex coefficients. An 8 × 8 system therefore has 64 complex channel coefficients to represent. Estimating these coefficients from noisy observations becomes increasingly challenging as the system dimension increases. The larger number of coefficients motivates investigation of learned channel-estimation methods. Nevertheless, demonstrating that a model generalises to larger antenna configurations requires direct experiments across different system dimensions; performance in the 8 × 8 simulation alone does not establish generalisation to all massive MIMO systems.

I. Computational Complexity of Detection

The computational requirements of a detection algorithm are an important consideration in multi-antenna communication. Linear detectors typically involve matrix operations, including matrix multiplication and, depending on the implementation, matrix inversion or pseudoinversion. The computational burden generally increases with the number of antennas. For a square matrix of dimension (N), direct matrix inversion has a conventional computational complexity on the order of (O(N^3)). This provides a useful broad estimate, although the actual cost depends on matrix

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

structure, numerical method, hardware, and implementation. Maximum-likelihood detection instead searches for the candidate symbol vector that minimises the distance between the received signal and the signal predicted by the channel model. It can be expressed as

$$\hat{\mathbf{s}}_\mathrm{ML} = \arg\min_\{\mathbf{s} \in \mathcal{A}^{\mathrm{Nt}}\} \|\mathbf{y} - \mathbf{Hs}\|^2_2 \quad (13)$$

where A is the modulation constellation and $\|\cdot\|_2$ denotes the Euclidean norm.For 16-QAM and eight transmitted streams, the candidate set contains

$$|\mathcal{A}|^{\mathrm{Nt}} = 16^8 = 4{,}294{,}967{,}296 \quad (14)$$

possible symbol vectors. An exhaustive search across this set for each received vector is generally impractical for the simulation's intended setting. Linear detection therefore provides a computationally simpler reference. ZF can suppress inter-stream interference, but it may amplify noise and channel-estimation errors. MMSE incorporates a regularisation term associated with the noise level. These differences are relevant when interpreting the error-rate results. The current study focuses on the reported detection and estimation metrics. Unless runtime, memory consumption, or operation counts are measured and reported, it should not claim that the attention-based model has lower computational complexity than the conventional methods.

J. Relationship Between BER, SER, and Detection Accuracy

BER and SER measure different aspects of receiver performance. BER is the fraction of transmitted bits recovered incorrectly, whereas SER is the fraction of transmitted modulation symbols detected incorrectly. Symbol detection accuracy is the proportion of symbols detected correctly.

For a dataset containing (N_b) transmitted bits and (N_e) bit errors, BER is

$$\mathrm{BER} = \mathrm{N_e}/\mathrm{N_b} \quad (15)$$

Similarly, if N_s symbols are transmitted and N_se symbols are detected incorrectly, SER is

$$\mathrm{SER} = \mathrm{N_se}/\mathrm{N_s} \quad (16)$$

When symbol accuracy is defined as the fraction of symbols detected correctly, it is related to SER by

$$\mathrm{A_symbol} = 1 - \mathrm{SER} \quad (17)$$

For 16-QAM, each symbol represents four bits. A symbol error can therefore involve between one and four incorrect bits. Under consistent bit and symbol counting, this gives the bounds

$$\mathrm{BER} \leq \mathrm{SER} \leq 4\ \mathrm{BER} \quad (18)$$

These metrics provide complementary information. BER directly reflects bit-recovery errors, while SER indicates how frequently the receiver selects an incorrect constellation point. Accuracy expresses the complementary proportion of correctly detected symbols and should not be interpreted as an independent performance measure when it is calculated directly from SER.

K. Effect of Channel-Estimation Error on ZF Detection

The quality of a channel estimate can influence the performance of a ZF detector. Let the estimated channel be expressed as

$$\hat{\mathbf{H}} = \mathbf{H} + \mathbf{E} \quad (19)$$

where E represents the channel-estimation error.Using the estimated channel, the ZF detector produces

$$\hat{\mathbf{x}} = \hat{\mathbf{H}}\dagger\mathbf{y} \quad (20)$$

Substituting the received-signal model y = Hx + n gives

$$\hat{\mathbf{x}} = \hat{\mathbf{H}}\dagger\mathbf{H}\mathbf{x} + \hat{\mathbf{H}}\dagger\mathbf{n} \quad (21)$$

This expression illustrates two important sources of detection error: imperfect channel knowledge and receiver noise. Even if the noise power decreases as the SNR increases, the effect of channel-estimation error may remain significant if the estimation error does not decrease correspondingly. Consequently, a learned estimator may experience an error floor if its predictions retain systematic or approximation errors at higher SNRs. Such behaviour is one possible explanation for a flattening error-rate curve, but it should not be treated as a proven cause without additional diagnostic experiments. In the reported results, the attention-based model's BER becomes relatively flat at higher SNRs and does not outperform the conventional MMSE baseline. This observation motivates further investigation of the estimator's prediction error, network capacity, training procedure, and interaction with ZF detection. These possibilities should be tested rather than presented as confirmed explanations.

L. Layers and Activation Functions

Neural networks use layers to transform input features into representations suitable for a prediction task. Fully connected layers learn weighted combinations of their input features, while nonlinear activation functions allow the model to represent relationships that cannot be expressed by a single linear transformation.

The rectified linear unit (ReLU) activation is defined as

$$\mathrm{ReLU}(z) = \max(0, z) \quad (22)$$

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

It retains positive input values and maps negative input values to zero. Other activation functions may be used depending on the model's design and output requirements. For channel estimation, the output must represent the real and imaginary components of the estimated channel coefficients, or an equivalent complex-valued representation supported by the implementation. The exact output arrangement must be stated clearly to explain how the network's predictions are converted into a channel matrix for detection. The general properties of fully connected layers and activation functions do not establish which specific configuration was used by the attention model in this study. The number of layers, units per layer, activation functions, output dimensions, and attention configuration must be reported from the actual model definition in Section IV.

M. Training by Gradient-Based Optimisation

A neural network is trained by adjusting its parameters to reduce a defined loss function.

$$\theta_t = \theta_\{t-1\} - \eta\nabla_\theta\mathcal{L}(\theta) \qquad (23)$$

where η is the learning rate and the gradient indicates the direction of greatest local increase in the loss.The Adam optimiser is a commonly used adaptive optimisation method that maintains moving estimates of the gradients and their squared values. Its parameter updates use these estimates to adjust the effective step size for each trainable parameter. Training duration, batch size, learning rate, loss function, and data partitioning can all influence the performance of a learned estimator. The manuscript reports 3,000 training samples, 500 testing samples, and 15 training epochs. These values describe the reported experiment, but they are not sufficient on their own to reproduce the training procedure. The learning rate, exact loss function, validation strategy, random seed, and model configuration should be added only after they have been verified against the implementation. If a separate validation set was not used, the paper must not imply that validation loss was monitored or that early stopping was applied. Similarly, the training and testing samples should be described in sufficient detail to establish how the reported performance was obtained.

N. Benefits and Challenges of Massive MIMO

Massive MIMO can improve wireless communication by exploiting spatial multiplexing, array gain, and spatial diversity. Spatial multiplexing allows multiple data streams to use the same time-frequency resources, while array processing can improve the effective signal quality. Multiple antenna paths can also provide diversity against fading and help the receiver manage interference. These benefits come with practical challenges. A larger antenna configuration increases the number of channel coefficients that must be estimated and can increase the computational requirements of detection. Pilot transmission consumes resources that could otherwise carry data, and non-orthogonal pilot signals may introduce pilot contamination in multi-user environments. Reliable channel estimation is therefore an important part of the receiver design. An inaccurate channel estimate can degrade subsequent signal detection, while a computationally demanding estimator may create implementation challenges. The design problem involves balancing estimation accuracy, detection reliability, computational cost, and the resources required to train or operate a learned model. This study investigates one aspect of that problem by evaluating an attention-based channel-refinement approach in an 8 × 8 Rayleigh-fading MIMO simulation with 16-QAM and AWGN. The reported results show improvement over the FFNN baseline under the evaluated conditions, but they do not establish that the attention-based method is superior to MMSE across the tested SNR range. Further experiments are required to establish its robustness, generalisation, and practical value.

III. LITERATURE REVIEW

A. Deep Learning for MIMO Channel Estimation and Signal Detection

Deep learning has gained attention in wireless communication because neural networks can learn complex relationships between transmitted signals, channel conditions, and received observations. In Massive Multiple Input Multiple Output (MIMO) systems, these capabilities can support channel estimation and signal detection, particularly when conventional methods face limitations caused by noise, interference, and increasing system dimensions. Babu et al. [1] proposed a pilot-assisted deep-learning approach for channel estimation and signal detection under a zero-channel-state-information (Zero-CSI) setting. Their method uses two feed-forward neural networks, one for channel estimation and another for signal detection. Although the approach avoids requiring prior channel-state information at the receiver, its performance remains below that of the CSI-based minimum mean-square-error (MMSE) benchmark in the reported experiments. Wang et al. [2] investigated pilot-assisted channel estimation and signal detection in uplink multi-user MIMO systems using deep learning. Their work demonstrates the potential of data-driven methods for improving communication performance. However, the effectiveness of such approaches depends on the availability and representativeness of the training data. Khani et al. [3] proposed adaptive neural signal detection for Massive

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

MIMO systems. Their work explored the use of learning-based detection to improve detection performance while considering computational complexity. These studies collectively indicate that neural networks can provide useful alternatives to conventional signal-processing techniques, although their effectiveness depends on model design, training conditions, and the evaluation environment.

B. AI-Native 6G Air Interface

The development of sixth-generation (6G) wireless networks has increased interest in integrating artificial intelligence into physical-layer communication processes. Hoydis et al. [4] discussed the concept of an AI-native air interface, in which learning-based techniques can assist or replace selected conventional communication functions. Such integration offers opportunities to adapt signal processing to complex wireless environments. Nevertheless, practical deployment requires careful consideration of computational cost, training requirements, reliability, and generalisation to channel conditions that may differ from those encountered during training. Babu et al. [5] proposed AIDETECT2, an AI-driven signal-detection approach for beyond-5G and 6G networks. Daha et al. [6] compared AI-based detection approaches with approximate and linear detection methods. These studies provide additional context for evaluating learned detectors against established signal-processing techniques rather than considering neural-network performance in isolation. Albreem et al. [9] reviewed deep-learning-based uplink detectors for Massive MIMO and discussed their computational requirements and unresolved research challenges. Their review highlights the importance of balancing detection performance with implementation complexity when considering learning-based methods for large antenna systems.

C. Attention-Based Channel Estimation

Attention mechanisms provide a way for a neural network to assign different weights to elements of its input according to their relevance to the task. In channel estimation, this mechanism may help a model identify informative relationships within noisy channel observations and learn features that are difficult to represent using a conventional feed-forward structure. Gao et al. [7] investigated an attention-aided deep-learning framework for Massive MIMO channel estimation and reported improvements in channel-estimation performance with a relatively small increase in complexity. Their work provides a motivation for examining attention mechanisms in learning-based channel estimation. Xiao et al. [10] studied multi-scale attention for channel estimation in reconfigurable intelligent surface (RIS)-aided Massive MIMO systems. Their approach explores attention at multiple feature scales within a specialised wireless configuration. Although this setting differs from the conventional Rayleigh-fading MIMO model considered in the present study, it illustrates the broader use of attention-based architectures in channel-estimation research. Gao et al. [11] investigated deep-learning-based channel estimation under channel aging and pilot contamination. These conditions introduce additional challenges because the channel may change over time and pilot observations may be affected by interference. Their work provides context for studying the robustness of learned estimation methods under non-ideal channel conditions. The findings of these studies motivate further investigation of attention-based channel estimation. However, improvements reported in other architectures or channel environments cannot automatically be assumed to apply to the model evaluated in this paper. Performance must be established through the experiments conducted under the present study's specific conditions.

D. Research Gap

Existing research has explored conventional estimation, feed-forward neural networks, model-driven learning, and attention-based channel estimation. Despite this progress, the relative performance of these approaches depends on the antenna configuration, channel model, training-data size, network architecture, and detection method. The base paper [1] provides a relevant starting point because it investigates pilot-assisted channel estimation and signal detection without prior CSI. However, its feed-forward architecture exhibits a performance gap relative to the CSI-based MMSE benchmark in the reported experiments. Attention-based channel estimation has shown potential in related studies [7], but the results of those studies do not establish how an attention-based estimator will perform when applied to the specific 8×8 MIMO configuration considered here. Accordingly, this work examines the replacement of the feed-forward channel-estimation model with an attention-based model that refines an initial least-squares (LS) channel estimate. The evaluation compares the resulting approach with LS, MMSE, and the feed-forward neural-network (FFNN) baseline under the implemented simulation conditions. The research gap is therefore focused on the comparative behaviour of these methods in the selected experimental setting. The objective is not to assume that attention will outperform every conventional estimator, but to determine where it improves performance and where limitations remain.

E. Comparison of Existing Approaches

The reviewed literature can be organised into three broad categories. The first category consists of data-driven neural-network methods for channel estimation or signal detection

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

[1]–[3], [5], [6]. These methods learn relationships from training examples and may capture nonlinear characteristics that are difficult to represent through conventional linear processing. Their effectiveness, however, depends on the training data, model architecture, and similarity between training and evaluation conditions. The second category consists of model-driven approaches that combine established signal-processing algorithms with learned components [8]. By incorporating known system structure into the learning process, these methods seek to balance model flexibility with computational efficiency and training-data requirements. The third category includes attention-based methods for channel estimation [7], [10], [11]. These studies investigate attention in conventional Massive MIMO, RIS-aided systems, and environments affected by channel aging or pilot contamination. Their results provide motivation for examining attention-based estimation, although their specific configurations and assumptions differ from those used in this study.

Table I. Table Of Related Work

Ref.	Approach	Main contribution or consideration
[1]	Pilot-assisted FFNN under Zero-CSI	Estimates the channel and detects data without prior CSI; reports a performance gap relative to CSI-based MMSE
[2]	Deep learning for multi-user uplink MIMO	Investigates learning-based estimation and detection; training-data requirements are an important consideration
[3]	Adaptive neural signal detection	Studies learning-based detection for Massive MIMO
[4]	AI-native 6G air interface	Discusses the integration of AI into future wireless communication systems
[5]	AIDETECT2	Investigates AI-driven detection for beyond-5G and 6G networks
[6]	AI, approximate, and linear detector comparison	Examines differences between learned and conventional detection approaches
[7]	Model-driven deep learning	Combines signal-processing structure with learning-based methods
[8]	Survey of deep-learning uplink detectors	Reviews complexity and open research challenges
[9]	Multi-scale attention for RIS-aided MIMO	Studies attention-based channel estimation in RIS-aided systems
[10]	Learning-based estimation under channel aging and pilot contamination	Investigates estimation in time-varying and interference-affected environments

The comparison indicates that no single method should be assumed to be optimal under all conditions. Conventional linear techniques provide important reference points, while learned approaches offer flexibility at the cost of additional architectural and training considerations.

F. Lessons from the Literature

Four observations from the reviewed studies guide the present work. First, pilot-assisted LS estimation provides a direct initial estimate of the channel from known pilot symbols. Using this estimate as the input to a learned refinement model provides a clear way to investigate whether learning can improve upon the initial estimate. Second, attention-based channel estimation has demonstrated potential in related Massive MIMO research [7]. This motivates evaluating attention as an alternative to the FFNN-based estimation approach used in the base paper. Third, the quantity and quality of training data are important considerations for data-driven methods [2]. Consequently, the limited training configuration used in the present work must be considered when interpreting its results. Fourth, conventional methods should remain part of the evaluation. The base paper [1] reports its performance relative to a CSI-based MMSE benchmark, and the present study retains this comparison to show whether the proposed attention-based model approaches that reference or remains behind it. These observations inform the model design and experimental comparisons. They are motivations derived from previous research, not experimental findings of the present study.

G. Positioning of the Present Work

This paper investigates a specific modification to a pilot-assisted MIMO estimation and detection pipeline. Instead of using the FFNN-based channel estimator from the base paper [1], the proposed approach uses an attention-based model to refine the initial LS estimate. The refined estimate is then supplied to a zero-forcing (ZF) detector. The evaluation uses an 8×8 MIMO configuration and compares four approaches: LS, MMSE, FFNN, and the attention-based model. Bit error rate (BER), symbol error rate (SER), and channel-estimation mean-square error (MSE) are used to examine performance across the evaluated signal-to-noise ratio (SNR) range. The study does not presume that the attention-based approach will outperform MMSE or LS at every SNR. Instead, it evaluates the observed differences between the methods and identifies the conditions under which the attention-based model performs favourably or unfavourably. This comparison is intended to provide a baseline for subsequent improvements to the model and its training procedure.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

H. Existing Methods and Their Drawbacks

LS estimation is straightforward to implement and does not require prior channel statistics. However, the estimate is affected by noise in the pilot observations, which can reduce its accuracy under noisy conditions. MMSE-based estimation or detection incorporates statistical information about the signal and noise to reduce estimation or detection error under its assumed model. Its implementation requires additional information and matrix operations, which can increase computational cost as the system dimension grows. Learning-based approaches offer an alternative by learning mappings from training examples. The base paper [1] uses separate feed-forward neural networks for channel estimation and signal detection without requiring prior CSI. Nevertheless, its reported comparison with CSI-based MMSE reveals that a learned approach does not necessarily achieve the performance of a conventional method supplied with channel information. Attention-based models may represent relationships within their input more flexibly than a plain feed-forward architecture. However, their performance still depends on the selected architecture, the training procedure, and the available data. Attention alone does not guarantee lower BER, SER, or channel-estimation MSE. These considerations motivate the present comparison of conventional and learned methods. They also emphasise the need to interpret the results in light of both the observed error rates and the limitations of the experimental configuration.

I. The Base Paper in Detail

The base paper [1] uses two feed-forward neural networks: one for channel estimation and another for signal detection. Each network contains three fully connected layers with clipped ReLU activation and a regression output layer. The reported training procedure uses the Adam optimiser with a learning rate of 0.001 for up to 2000 epochs. The study uses 0.5 million pilot symbols for training and tests on 1 million symbols. The base paper evaluates 2 × 2 and 4 × 4 antenna configurations using 16-QAM modulation over an SNR range from 0 to 20 dB in increments of 5 dB. Its inputs include pilot-symbol features and the LS channel estimate. The reported results are compared with CSI-based MMSE detection. The paper reports performance differences relative to MMSE of approximately 2% and 16% at low SNR for the 2 × 2 and 4 × 4 configurations, respectively. At high SNR, the reported differences increase to approximately 25% and 28%. The authors associate the larger performance gap in the 4 × 4 configuration with the increased number of variables and interference between antennas. Two important differences distinguish the present study from the base paper. First, the current approach replaces the learned detection network with a ZF detector so that the effect of the channel estimate can be evaluated through a common detection stage. Second, the present study uses 3000 training samples and 500 test samples over 15 training epochs, rather than the much larger training and testing configuration reported in [1]. These differences limit direct comparisons between the two studies. The sample definitions and training conditions may not be identical, and the architecture and detection pipeline are also different. Therefore, the results of the present work should be interpreted as findings from its own simulation configuration rather than as a direct reproduction of the base paper.

J. Notes on Individual Works

Wang et al. [2] investigated pilot-assisted channel estimation and signal detection in multi-user uplink MIMO systems. Their work highlights the role of deep learning in modelling wireless communication behaviour and the importance of sufficient training data. Khani et al. [3] studied adaptive neural detection for Massive MIMO. Their work contributes to the investigation of learning-based detectors and the trade-off between detection performance and implementation complexity. Hoydis et al. [4] discussed the integration of AI into the 6G air interface. This provides a broader system-level motivation for investigating learning-based physical-layer processing. Babu et al. [5] proposed AIDETECT2 for beyond-5G and 6G signal detection, while Daha et al. [6] compared AI-based methods with approximate and linear detectors. These studies support the importance of evaluating learned methods against established baselines. Gao et al. [7] examined attention-aided deep learning for Massive MIMO channel estimation. Their work directly motivates the use of attention in the present study, although its results do not predetermine the performance of the architecture evaluated here. He et al. [8] investigated model-driven deep learning for MIMO detection, demonstrating an alternative approach in which known signal-processing structure is incorporated into learning-based methods. Albreem et al. [9] reviewed deep-learning methods for Massive MIMO uplink detection and discussed associated complexity and open problems. Xiao et al. [10] studied multi-scale attention for channel estimation in RIS-aided Massive MIMO systems, while Gao et al. [11] investigated learning-based channel estimation under channel aging and pilot contamination. Both works illustrate the application of learned estimation methods to specialised wireless conditions. Overall, the literature supports further investigation of attention-based channel estimation while also showing why careful experimental comparison is necessary. The following section describes

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

the proposed methodology and the simulation configuration used to evaluate the attention-based estimator.

IV. PROPOSED METHODOLOGY

A. System Overview

The proposed system investigates attention-based channel estimation for an 8 × 8 Massive MIMO communication link. The transmitter generates random binary data and maps the bits to 16-QAM symbols. Known pilot symbols are used to obtain an initial channel estimate at the receiver. The transmitted signal passes through a Rayleigh fading channel, and additive white Gaussian noise (AWGN) is introduced according to the selected signal-to-noise ratio (SNR). At the receiver, the received pilot observations are used to calculate the least-squares (LS) channel estimate. This estimate is supplied to the attention-based neural network, which produces a refined channel estimate. The refined estimate is then used by the zero-forcing (ZF) detector to recover the transmitted symbols. The detected symbols are demodulated into bits, and the resulting bit and symbol errors are measured. The performance of the proposed method is compared with conventional LS and MMSE methods and the feed-forward neural network (FFNN) baseline. The evaluation considers bit error rate (BER), symbol error rate (SER), and channel-estimation mean-square error (MSE). Figure 1 presents the overall processing sequence, from transmission and channel propagation to estimation, detection, and performance evaluation.

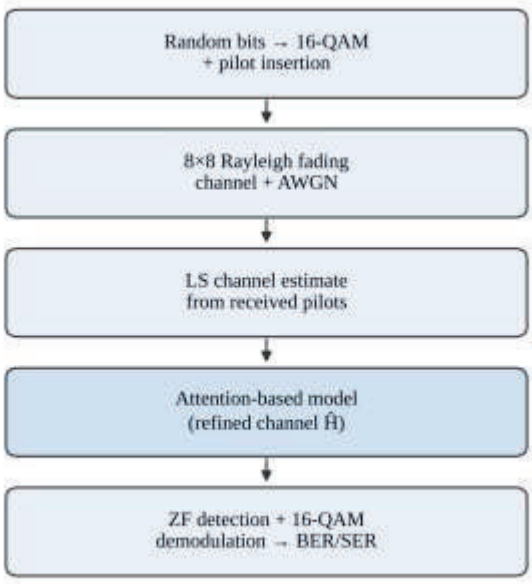


Fig. 1. Block diagram of the proposed system.

B. Transmitter, Channel, and Receiver

The transmitter generates random bits and groups them into four-bit sequences for 16-QAM modulation. The resulting constellation symbols are arranged into transmit streams and passed through the simulated MIMO channel. Pilot symbols are inserted to support channel estimation at the receiver. For the 8 × 8 configuration, the channel connects eight transmit antennas to eight receive antennas. Each transmit–receive antenna pair is represented by a complex channel coefficient following the Rayleigh fading model. AWGN is added to the received signal at the selected SNR. At the receiver, the received observations are processed to separate the pilot and data components. The pilot observations are used to compute the LS channel estimate, while the received data symbols are subsequently processed using the selected channel estimate and detector. The complex-valued LS estimate is represented using its real and imaginary components before being supplied to the neural network. The corresponding simulated channel provides the target during supervised training. This representation enables the network to process the channel information using real-valued neural-network operations.

C. Dataset and Simulation Setup

The dataset is generated using a Python-based simulation of an 8 × 8 MIMO system. Each training example contains an LS channel estimate obtained from noisy pilot observations, represented through its real and imaginary components. The target is the corresponding simulated channel matrix in the same representation. The simulation uses 16-QAM modulation, Rayleigh fading, and AWGN. The SNR is evaluated over the range from 0 to 20 dB. Transmitted symbols, received observations, channel estimates, and detected outputs are generated to support the evaluation of the competing methods.

Table II. Simulation Parameters

Parameter	Configuration
MIMO system	8 × 8
Modulation	16-QAM
Channel model	Rayleigh fading
Noise	AWGN
SNR range	0–20 dB
Training samples	3000
Testing samples	500
Epochs	15
Batch size	64
Optimiser	Adam
Loss function	Mean squared error (MSE)

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

Parameter	Configuration
Detector	ZF

The available experimental documentation does not fully specify what constitutes one sample, the number of pilot symbols, the SNR increment, or the number of Monte Carlo trials. These details should be obtained from the implementation before the experiment is considered fully reproducible. They should not be inferred from the number of training samples alone.

D. Attention-Based Channel Estimation Model

The proposed model receives the LS channel estimate and generates a refined estimate. Its input–output relationship is represented as

$$\mathbf{H}_\text{Att} = \text{f_Att}(\mathbf{H}_\text{LS}) \qquad (24)$$

where Ĥ_LS denotes the LS channel estimate, Ĥ_Att denotes the refined channel estimate, and f_Att represents the trained attention-based network. The network uses attention-based processing followed by fully connected layers to learn a mapping from the initial channel estimate to the corresponding channel target. During training, the simulated channel serves as the reference output. The objective is to reduce the difference between the predicted channel and the target channel. The exact architecture must be documented using the actual model implementation. The number and type of attention blocks, number of attention heads, layer dimensions, activation functions, input and output tensor shapes, and total number of trainable parameters should be reported from the code. These details are not sufficiently established by the available manuscript text and therefore must not be guessed.

E. Training Procedure

The attention-based network is trained using the LS channel estimate as input and the corresponding simulated channel as the target. The model parameters are updated to reduce the mean-squared difference between the estimated and reference channel coefficients.

For a batch of training examples, the loss is represented as

$$\mathcal{L} = (1/B)\, \Sigma_{i=1}{}^{B} \, \|\mathbf{H}_i - \mathbf{H}_\text{Att},_i\|^2_\text{F} \qquad (25)$$

where B is the batch size, H_i is the reference channel matrix for sample i, Ĥ_Att,i is the corresponding attention-based channel estimate, and ‖·‖_F denotes the Frobenius norm. The documented training configuration uses 3000 training samples, a batch size of 64, and 15 epochs. The Adam optimiser is used to update the trainable parameters. Training and validation losses are reported as being monitored during the training process; the actual loss curves should be inspected before drawing conclusions about convergence or overfitting. The limited number of samples and training epochs is an important consideration when interpreting the results. A model trained under these conditions may not learn all relevant channel characteristics, and additional runs would be needed to assess the stability of its performance.

F. Detection and Algorithm

After training, the attention-based model processes the LS estimate to obtain the refined channel estimate. This estimate is supplied to the ZF detector, using the ZF expression defined in Section II with the refined channel substituted for the true channel matrix. The detector estimates the transmitted symbols from the received signal and the refined channel estimate. The detected symbols are then mapped back to their corresponding bit sequences using 16-QAM demodulation. The detected bits and symbols are compared with the transmitted data to calculate BER and SER. Channel-estimation MSE is calculated by comparing the estimated channel with the simulated reference channel.

Algorithm 1. Attention-based channel estimation and detection

1. Generate random bits and map them to 16-QAM symbols.
2. Insert known pilot symbols and transmit the signal through the 8 × 8 Rayleigh fading channel.
3. Add AWGN according to the selected SNR.
4. Obtain the LS channel estimate from the received pilot observations.
5. Train the attention-based network using the LS estimate as input and the simulated channel as the target.
6. Apply the trained model to obtain the refined channel estimate.
7. Use the refined channel estimate in the ZF detector.
8. Demodulate the detected symbols into bits.
9. Calculate BER, SER, and channel-estimation MSE.
10. Repeat the evaluation for the required SNR values and compare the results with the baseline methods.

G. Performance Measures

Three measures are used to evaluate the proposed approach: BER, SER, and channel-estimation MSE. BER quantifies the fraction of transmitted bits detected incorrectly, while SER measures the fraction of transmitted symbols detected incorrectly. These measures describe the end-to-end

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

detection performance of the system. Channel-estimation MSE measures the difference between the reference channel matrix and its estimate. For a system with M receive antennas and N transmit antennas, the channel-estimation MSE is expressed as

$$\text{MSE_H} = (1/\text{MN})\ \mathbb{E}[\|\mathbf{H} - \hat{\mathbf{H}}\|^2\text{_F}] \quad (26)$$

where H is the reference channel matrix, $\hat{H}$ is the estimated channel matrix, and E[·] denotes the expectation over the evaluated channel samples. BER and SER are used to compare the detection methods, while channel-estimation MSE assesses the quality of the channel estimates themselves. Considering these measures together helps distinguish improvements in channel estimation from changes in the final detection performance.

H. Design Choices and Alternatives

The LS estimate is used as the input to the attention-based network rather than supplying the raw received pilot observations directly. This design allows the network to refine an initial channel estimate instead of learning the entire estimation process from the received observations. It also provides a consistent connection with the pilot-assisted estimation approach described in the base paper [1]. The MSE loss is selected because it directly measures the difference between the estimated and reference channel matrices. Adam is used as the optimiser for training the network. The ZF detector is selected so that the refined channel estimate can be used directly in the detection stage without introducing a second learned detection network. An alternative would be to use the refined channel estimate with an MMSE detector. Because MMSE includes a noise-dependent regularisation term, its response to channel-estimation errors can differ from that of ZF. This alternative is not established as part of the reported experiment unless it is implemented and evaluated separately. The selected design choices should therefore be understood as the configuration of this study rather than proof that they are universally optimal.

I. Computational Considerations

The LS estimator and the ZF detector involve matrix operations whose computational requirements increase with the number of antennas. The attention-based model introduces additional computational costs that depend on its architecture, input representation, layer dimensions, and attention configuration. The available manuscript does not provide a verified operation count, parameter count, memory measurement, or execution-time comparison for the implemented attention model. Consequently, a quantitative claim that the proposed method is computationally cheaper or more expensive than LS, MMSE, or FFNN cannot be made from the documented information alone. A complete complexity analysis should report the number of trainable parameters and, where feasible, inference time and computational operations under a common hardware and software configuration. Until those measurements are available, the comparison should focus on the measured estimation and detection performance.

J. Implementation and Baselines

The simulation is implemented in Python using TensorFlow/Keras for the neural network, NumPy for numerical operations, and Matplotlib for result visualisation. The documented development environments include Google Colab and Jupyter Notebook. The proposed attention-based method is compared with LS, MMSE, and the FFNN baseline. The comparison is intended to show how the refined channel estimate affects performance relative to conventional estimation and the existing learned approach.

Table III explains the principal differences between the base paper and the present implementation.

Table III. Comparison With The Base Paper [1]

Aspect	Base paper [1]	Present study
Learning model	FFNN-based approach	Attention-based channel estimator
Initial channel estimate	LS estimate from pilots	LS estimate from pilots
Detection stage	Learned detection network	ZF detector using the refined estimate
Antenna configuration	2×2 and 4×4	8×8
Modulation	16-QAM	16-QAM
Baselines	CSI-based MMSE	LS, MMSE and FFNN

The two studies differ in antenna configuration, model structure, detection method, and training scale. Therefore, differences between their reported results cannot be attributed solely to the use of attention. The present work should be evaluated primarily against its own baselines under the documented simulation conditions.

K. Reproducibility

The available documentation specifies the antenna configuration, modulation, channel and noise models, SNR range, training and testing sample counts, batch size, number of epochs, optimiser, and loss function. However, several details needed for exact reproduction remain unspecified, including the random seed, number of pilot symbols, definition of a sample, SNR increment, number of Monte Carlo trials, and full attention-network architecture.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

These missing details can influence the resulting error rates and training behaviour. The relatively small training set and limited number of epochs may also cause results to vary across different random initialisations or simulation runs. For stronger reproducibility, the implementation should record all simulation parameters, model settings, software versions, and random seeds. Repeating the experiment with multiple seeds and reporting the mean and variability of the measured metrics would provide a more reliable assessment of the proposed method.

L. Walk-Through for One Sample

For an individual sample, random information bits are first mapped to 16-QAM symbols and assigned to the transmit streams. Known pilot symbols are transmitted through the simulated MIMO channel, and AWGN is added according to the selected SNR. At the receiver, the pilot observations are extracted and used to calculate the LS channel estimate. Its real and imaginary components are provided to the trained attention-based model, which produces a refined channel estimate. The received data signal is then processed by the ZF detector using this refined estimate. The detected constellation symbols are mapped back to bits. The resulting bit and symbol sequences are compared with the original transmitted data to determine the detection errors. The estimated channel is also compared with the simulated reference channel to calculate channel-estimation MSE. Repeating this process over the test samples provides the measurements used in the results section.

M. Evaluation Protocol

The evaluation compares the LS, MMSE, FFNN, and attention-based approaches over the selected SNR range. The test samples are kept separate from the training data so that the final measurements assess performance on samples not used for parameter optimisation. For a fair comparison, the methods should be evaluated using consistent transmitted data, channel realisations, noise conditions, and metric definitions wherever their implementations permit. If different random data or channel realisations are used, the resulting variation should be considered when interpreting differences between methods. The available documentation does not fully establish every detail of how the baseline implementations share data and noise. This procedure should therefore be checked against the simulation code before claiming that all methods were evaluated under identical realisations. Similarly, any comparison involving MMSE must specify how that baseline is implemented, including the channel information and noise variance supplied to it. The BER, SER, and channel-estimation MSE values are calculated at the evaluated SNR points and presented in Section V. Conclusions are restricted to the tested conditions and should not be interpreted as evidence of performance at untested SNR values or in other channel environments.

N. Software and Hardware

The documented software environment uses Python 3.11 or later, TensorFlow/Keras, NumPy, and Matplotlib, with Google Colab and Jupyter Notebook listed as development environments. The stated system requirements include an Intel Core i5/i7 or AMD Ryzen 5 or higher processor, at least 8 GB of RAM, and Windows 10/11 or Ubuntu Linux. The exact hardware configuration, library versions, execution environment used for the reported results, and total runtime are not established by the available documentation. These details should be added from the actual environment if reproducibility or computational efficiency is being assessed. Overall, the methodology defines a pilot-assisted MIMO simulation in which an attention-based model refines the LS channel estimate before ZF detection. The subsequent results evaluate the approach against the selected baselines using BER, SER, and channel-estimation MSE, while recognising the limitations imposed by the documented training scale and incomplete implementation details.

V. RESULTS AND DISCUSSION

This section evaluates the attention-based channel estimation model in an 8 × 8 Massive MIMO system using 16-QAM modulation, Rayleigh fading, and additive white Gaussian noise (AWGN). The performance is compared with the Least Squares (LS), Minimum Mean Square Error (MMSE), and Feed-Forward Neural Network (FFNN) methods at signal-to-noise ratio (SNR) values of 0, 5, 10, 15, and 20 dB. The evaluation considers bit error rate (BER), symbol error rate (SER), detection accuracy, and the variation in error rates with increasing SNR.

The reported results are based on the simulation values available for this study. Repeated runs, confidence intervals, and statistical significance tests are not reported; therefore, the findings should be interpreted as results from the evaluated simulation rather than as evidence of statistical generalisation.

A. BER Performance

BER measures the proportion of transmitted bits that are detected incorrectly. A lower BER indicates more reliable data transmission. Table IV presents the BER results obtained for the four methods over the evaluated SNR range.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

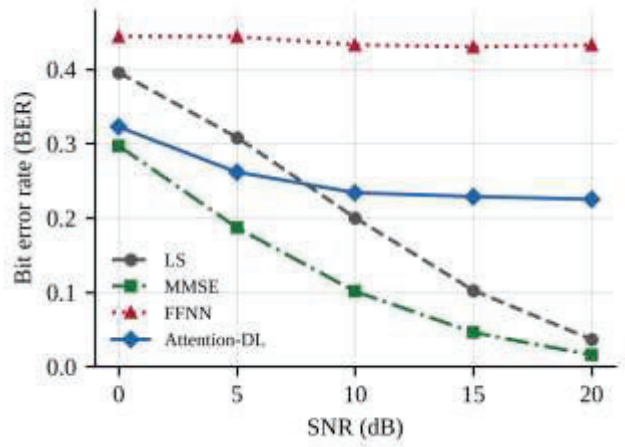


Fig. 2. BER versus SNR for LS, MMSE, FFNN and the attention model.

Table IV. Ber Versus Snr

SNR (dB)	LS	MMSE	FFNN	Attention
0	0.39569	0.29700	0.44406	0.32231
5	0.30838	0.18688	0.44381	0.26162
10	0.20019	0.10125	0.43288	0.23394
15	0.10269	0.04612	0.42994	0.22862
20	0.03669	0.01581	0.43219	0.22525

The results show that the MMSE method achieves the lowest BER at every evaluated SNR. Its BER decreases from 0.29700 at 0 dB to 0.01581 at 20 dB. The LS estimator also improves consistently, with BER decreasing from 0.39569 to 0.03669 over the same range. The attention-based model performs better than the FFNN baseline at all five SNR values. At 0 dB, its BER is 0.32231, compared with 0.44406 for the FFNN. At 20 dB, the corresponding values are 0.22525 and 0.43219. This represents an approximately 47.9% reduction in BER relative to the FFNN at 20 dB. The comparison with LS produces a different result. The attention model achieves lower BER at 0 and 5 dB, with values of 0.32231 and 0.26162, respectively. At 10, 15, and 20 dB, however, LS performs better. At 20 dB, the LS BER is 0.03669, substantially below the attention model's value of 0.22525. A notable observation is that the attention model's BER decreases relatively slowly above 10 dB. It changes from 0.23394 at 10 dB to 0.22525 at 20 dB, indicating a pronounced flattening of the error curve. Consequently, the model does not benefit from increasing SNR to the same extent as LS and MMSE in the evaluated configuration.

B. SER Performance

SER measures the proportion of transmitted symbols detected incorrectly. Since 16-QAM represents each symbol using four bits, a symbol error occurs whenever the detected symbol differs from the transmitted symbol, even if only one of its bits is incorrect.

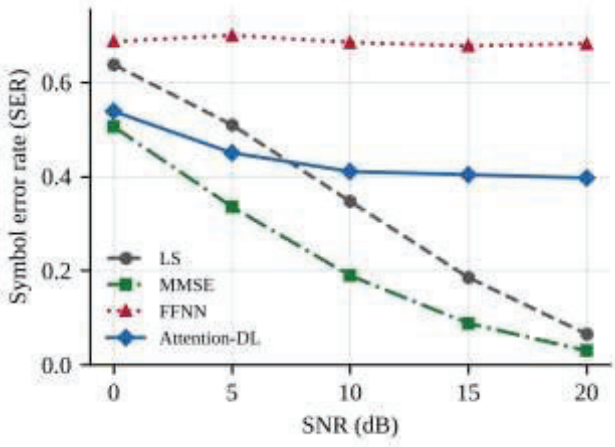


Fig. 3. SER versus SNR for LS, MMSE, FFNN and the attention model.

Table V. Ser Versus Snr

SNR (dB)	LS	MMSE	FFNN	Attention
0	0.63713	0.50513	0.68588	0.53775
5	0.50925	0.33537	0.69925	0.44988
10	0.34700	0.18925	0.68437	0.41025
15	0.18538	0.08825	0.67688	0.40375
20	0.06563	0.03000	0.68150	0.39663

The SER results broadly support the BER findings. MMSE achieves the lowest SER at every evaluated SNR, followed by LS at the higher SNR values. The attention model consistently achieves a lower SER than the FFNN, but it does not outperform MMSE at any tested SNR. At 20 dB, the attention model produces an SER of 0.39663, compared with 0.68150 for the FFNN. This corresponds to an approximately 41.8% reduction in SER. However, the attention model's SER remains substantially higher than the LS and MMSE values of 0.06563 and 0.03000, respectively. The attention model also shows limited improvement as SNR increases beyond 10 dB. Its SER changes from 0.41025 at 10 dB to 0.39663 at 20 dB. In comparison, the LS and MMSE methods exhibit much larger reductions in symbol errors across this range. These results indicate that the attention-based approach improves upon the evaluated FFNN baseline but does not yet provide the reliability

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

achieved by the conventional LS and MMSE methods at medium and high SNR.

C. Detection Accuracy

Detection accuracy represents the proportion of transmitted symbols detected correctly. In this study, it is derived from SER as one minus the symbol error rate. Table VI presents the resulting accuracy values as percentages. The accuracy results demonstrate that MMSE provides the highest detection accuracy throughout the tested SNR range. At 20 dB, its accuracy reaches 97.0%, while LS achieves 93.4%. The attention model reaches 60.3%, compared with 31.9% for the FFNN. The attention model's accuracy increases from 46.2% at 0 dB to 59.0% at 10 dB. It then rises by only 1.3 percentage points between 10 and 20 dB. This trend is consistent with the flattening observed in its BER and SER curves. The FFNN remains near 31% accuracy across the evaluated range. Although the attention model provides a substantial improvement over this baseline, its limited accuracy gain at higher SNR suggests that the present model configuration does not adequately exploit the improved channel conditions.

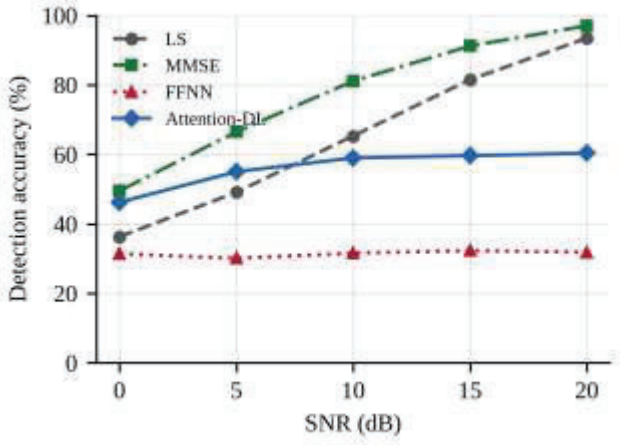


Fig. 4. Detection accuracy (1 − SER) versus SNR.

Table VI. Detection Accuracy (%) Versus Snr

SNR (dB)	LS	MMSE	FFNN	Attention
0	36.3	49.5	31.4	46.2
5	49.1	66.5	30.1	55.0
10	65.3	81.1	31.6	59.0
15	81.5	91.2	32.3	59.6
20	93.4	97.0	31.9	60.3

D. Relative Improvement of the Attention Model

To quantify the difference between the attention model and each comparison method, the relative change in error rate is calculated using the reference method's error rate as the denominator. A positive value indicates that the attention model has a lower error rate; a negative value indicates that its error rate is higher. The relative changes reported in Table VII are reproduced below.

Table VII. Relative Change Of The Attention Model (%)

SNR	BER vs FFNN	BER vs LS	BER vs MMSE	SER vs FFNN	SER vs LS	SER vs MMSE
0	27.4	18.5	−8.5	21.6	15.6	−6.5
5	41.1	15.2	−40.0	35.7	11.7	−34.1
10	46.0	−16.9	−131.1	40.1	−18.2	−116.8
15	46.8	−122.6	−395.7	40.4	−117.8	−357.5
20	47.9	−513.9	−1324.7	41.8	−504.3	−1222.1

The attention model's advantage over the FFNN increases with SNR. The relative BER reduction rises from 27.4% at 0 dB to 47.9% at 20 dB. Its relative SER reduction also increases, reaching 41.8% at 20 dB. Against LS, the attention model performs better at 0 and 5 dB but worse at 10 dB and above. The results therefore indicate a crossover between 5 and 10 dB. This suggests that the attention-based refinement is beneficial in the lower-SNR conditions represented in this simulation, but it does not maintain that advantage as the noise level decreases. The attention model performs worse than MMSE at all tested SNR values. The large negative relative changes at high SNR arise because the MMSE error rates become much smaller than those of the attention model. These figures should not be interpreted as percentage-point differences; they are relative changes calculated against the reference error rate.

E. Gap to MMSE and Crossover with LS

The relative BER values provide another way to examine the performance difference between the attention model and the conventional estimators. The ratio of the attention model's BER to the MMSE BER is approximately 1.09, 1.40, 2.31, 4.96, and 14.25 at 0, 5, 10, 15, and 20 dB, respectively. This increasing ratio shows that the gap between the attention model and MMSE widens as SNR increases. At 20 dB, the attention model's BER is approximately 14.25 times the MMSE BER. The ratio between the attention model's BER and the LS BER follows a different pattern. It is approximately 0.81 at 0 dB and 0.85 at 5 dB, indicating lower BER for the attention model at these two operating points. The ratio increases to approximately 1.17 at 10 dB, 2.23 at 15 dB, and 6.14 at 20 dB. This confirms that the attention and LS curves cross

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

between 5 and 10 dB. These comparisons highlight an important distinction: improvement over a neural-network baseline does not necessarily imply improvement over conventional channel estimation. In the present experiment, the attention model offers a low-SNR advantage over LS but loses that advantage at higher SNR.

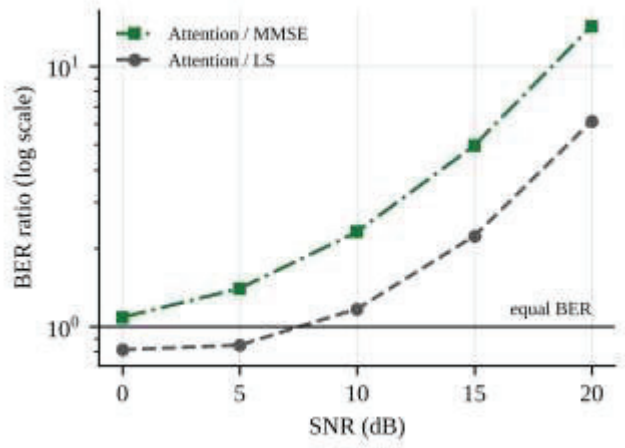


Fig. 5. BER of the attention model relative to MMSE and LS.

F. BER Reduction Across SNR Increments

The change in BER between successive SNR values provides insight into how effectively each method benefits from an improvement in channel conditions. For the attention model, the BER decreases by approximately 0.061 between 0 and 5 dB, by 0.028 between 5 and 10 dB, by 0.005 between 10 and 15 dB, and by 0.003 between 15 and 20 dB. For LS, the corresponding decreases are approximately 0.087, 0.108, 0.098, and 0.066. For MMSE, they are approximately 0.110, 0.086, 0.055, and 0.030. The FFNN shows only small changes, with its BER remaining close to 0.43. Across the complete 0–20 dB range, the attention model's BER decreases by 0.097. The LS BER decreases by 0.359, and the MMSE BER decreases by 0.281. The attention model therefore captures a smaller portion of the improvement observed in the conventional estimators, with most of its BER reduction occurring in the first 10 dB.

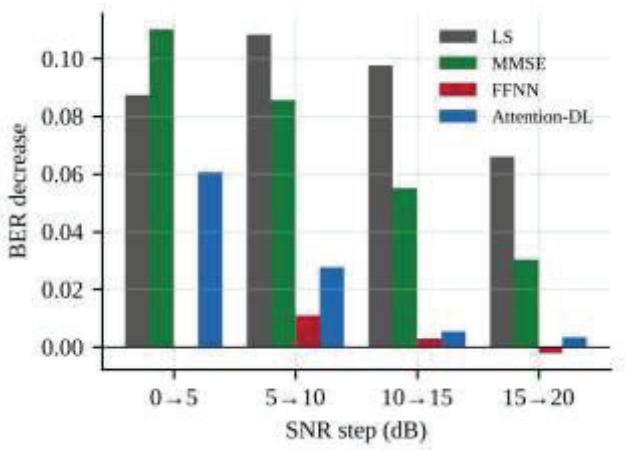


Fig. 6. Decrease of BER for each 5 dB step of SNR.

G. Structure of the Errors

The ratio of SER to BER offers a simple way to examine the relationship between symbol errors and bit errors. For 16-QAM, each erroneous symbol can contain between one and four incorrect bits. Consequently, the ratio varies according to the distribution of bit errors among the incorrectly detected symbols.

Table VIII. Ratio Ser/Ber

SNR (dB)	LS	MMSE	FFNN	Attention
0	1.61	1.70	1.54	1.67
5	1.65	1.79	1.58	1.72
10	1.73	1.87	1.58	1.75
15	1.81	1.91	1.57	1.77
20	1.79	1.90	1.58	1.76

The SER-to-BER ratios range from approximately 1.54 to 1.91 across the evaluated methods and SNR values. The attention model's ratio increases gradually from 1.67 to 1.77. The FFNN produces the lowest ratios, remaining between approximately 1.54 and 1.58. These results suggest that the main performance differences between the methods are associated with how frequently symbols are detected incorrectly, rather than with a major difference in the number of incorrect bits within each erroneous symbol. This interpretation is limited to the aggregate error-rate measurements available in the simulation; detailed error-pattern analysis would be needed to establish the underlying causes.

H. Overall Discussion

Three principal findings emerge from the results.

First, the attention-based model consistently improves upon the FFNN baseline. At 20 dB, the BER decreases from

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

0.43219 to 0.22525, while the SER decreases from 0.68150 to 0.39663. The detection accuracy also increases from 31.9% to 60.3%. This supports the use of the attention-based approach as a more effective alternative to the evaluated FFNN configuration. Second, the attention model improves upon LS only at the two lowest SNR values. Its relative BER reductions are 18.5% at 0 dB and 15.2% at 5 dB. At higher SNR values, LS performs better. The observed crossover suggests that the benefit of the learned refinement depends on the operating conditions and the current model configuration. Third, MMSE remains the strongest method among those evaluated. It achieves the lowest BER and SER at every SNR value. At 20 dB, its BER is 0.01581 and its SER is 0.03000, both substantially lower than the corresponding attention-model values. The attention model's relatively flat BER curve above 10 dB is an important limitation. The current results do not establish the exact cause of this behaviour. Potential explanations include limited training data, insufficient training epochs, an architecture that does not adequately represent the channel-estimation task, or sensitivity introduced by the subsequent Zero-Forcing (ZF) detector. These explanations remain hypotheses and require additional experiments.

I. Possible Reasons for the Flat BER Curve

Several factors may contribute to the limited high-SNR improvement observed in the attention model. Limited training data and duration: The reported setup uses 3,000 training samples and 15 epochs. These settings may be insufficient for learning a robust mapping from noisy LS channel estimates to accurate channel estimates. However, the available results alone cannot confirm that increasing the dataset or training duration will remove the observed error floor. Model architecture: The detailed attention configuration, including the number of attention blocks, number of heads, layer dimensions, activation functions, and trainable parameter count, must be documented. Without these details, it is difficult to determine whether the model has sufficient capacity or whether its implementation is appropriate for the input representation. Dependence on the channel conditions: The experiment considers an 8 × 8 Rayleigh fading channel with AWGN. The results do not establish whether the trained model generalises to other channel models, pilot configurations, antenna arrangements, or changing channel conditions. Interaction with the ZF detector: The attention model refines the LS channel estimate, after which ZF detection is performed. Errors in the refined channel estimate can affect the detected symbols, and ZF can amplify noise when the channel matrix is poorly conditioned. A comparison using both ZF and MMSE detection with the same refined estimate would help separate the contribution of channel estimation from that of the detection algorithm. SNR representation during training: The present results do not establish whether the model receives SNR as an explicit input or how training samples are distributed across SNR values. Further testing should examine whether the model can adapt its refinement to different noise levels.

These explanations should be treated as directions for investigation rather than verified causes. Additional experiments, including more training samples, longer training, multiple random seeds, detailed architecture reporting, and comparisons with alternative detectors, are needed to determine which factors most strongly influence performance. Overall, the results demonstrate a measurable improvement over the evaluated FFNN baseline, particularly in BER and SER. Nevertheless, the attention model's limited high-SNR improvement and its performance gap relative to MMSE indicate that further development is necessary before it can be considered a replacement for conventional estimation and detection methods.

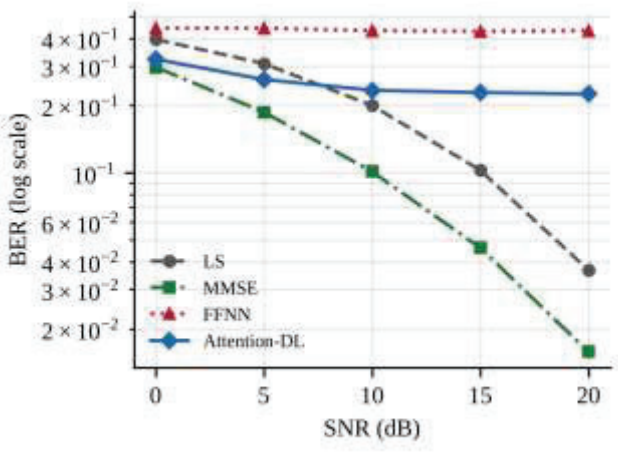


Fig. 7. BER versus SNR on a logarithmic axis.

Figure 7 presents the BER performance of the LS, MMSE, FFNN, and attention-based methods using a logarithmic vertical axis. This representation makes it easier to compare error rates across different magnitudes, particularly when the MMSE and LS methods achieve lower BER values than the neural-network-based methods at higher SNR levels. The figure highlights the consistent reduction in BER achieved by MMSE, the improvement of the attention model over the FFNN, and the attention model's limited reduction in BER above 10 dB. These observations reinforce the results presented in Table IV.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

VI. LIMITATIONS AND THREATS TO VALIDITY

Although the attention-based model demonstrates an improvement over the feed-forward neural network (FFNN) baseline, several limitations must be considered when interpreting the results. These limitations concern the size of the training dataset, the range of signal-to-noise ratio (SNR) values, computational requirements, generalisation to other channel conditions, and the validity of the selected performance measures.

A. Limited Training Data and Training Duration

The proposed model was trained using a synthetic dataset containing 3,000 training samples and evaluated using 500 testing samples over 15 training epochs. Although these settings allow an initial comparison between the attention-based model and the selected baselines, they provide limited evidence of the model's performance under a wider range of operating conditions. Deep-learning-based channel estimation methods depend on the quality, diversity, and quantity of training data. A small dataset may not represent the full range of channel variations encountered in practical wireless communication systems. Similarly, training for a limited number of epochs may not provide sufficient opportunities for the model to learn the underlying channel characteristics. Consequently, the observed performance should be interpreted as preliminary rather than as evidence that the proposed architecture has reached its best achievable performance. Future experiments should evaluate larger datasets, longer training schedules, and multiple independent training runs to determine whether the observed improvements remain consistent.

B. Limited SNR Range and Performance at High SNR

The experiments evaluate the system over an SNR range of 0–20 dB. Therefore, the reported results do not establish how the model performs outside this interval. Its behaviour under extremely noisy conditions or at higher SNR values remains unknown. A further limitation is the relatively flat BER curve of the attention-based model above approximately 10 dB. While the BER of the conventional LS estimator continues to decrease as the SNR increases, the attention-based model does not demonstrate a comparable reduction. At higher SNR values, LS consequently outperforms the attention-based approach in the reported experiments. This behaviour indicates that the current model does not consistently exploit the improved signal quality available at higher SNR values. The precise cause cannot be established from the reported results alone. Further investigation of the network architecture, training procedure, dataset distribution, and estimation errors is required to identify the factors responsible for this behaviour. Future work should evaluate more SNR values and investigate training strategies that improve performance across the complete operating range.

C. Data Availability and Training Cost

The present investigation uses simulated data generated under a Rayleigh fading channel model with additive white Gaussian noise (AWGN). Although simulation provides a controlled environment for comparing estimation methods, simulated samples may not capture all the characteristics of practical wireless channels. In a real communication system, training data may need to be obtained through measurements, a validated channel simulator, or a combination of both. Collecting representative data and retraining a model for different environments may introduce additional time, computational, and operational costs. The current experiments do not quantify the amount of data required to achieve a specified performance level or the cost of repeatedly training the model under changing channel conditions. Therefore, the practical data requirements of the proposed approach remain an open question. Future investigations should examine the relationship between dataset size, training duration, channel diversity, and estimation performance. The use of measured channel data should also be explored to assess whether the observed improvements extend beyond the simulated environment.

D. Computational Complexity and Processing Latency

Low latency is an important requirement for emerging 6G communication systems. Although attention-based learning provides a mechanism for processing informative input features, its practical suitability depends on the computational resources required for training and inference. The present evaluation focuses primarily on BER, SER, detection accuracy, and channel-estimation performance. It does not establish the model's inference latency, memory requirements, computational complexity, or energy consumption. Consequently, the results are insufficient to conclude that the proposed model satisfies the processing requirements of a practical 6G receiver. A meaningful comparison should measure the computational cost of the attention-based model against LS, MMSE, and FFNN-based alternatives under consistent hardware and evaluation conditions. Such measurements would help determine whether the observed improvement over the FFNN justifies the additional processing required by the learned estimator. Future work should include inference-time measurements, parameter counts, memory usage, and computational-cost analysis. These evaluations would provide a more complete

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

basis for assessing the model's suitability for real-time deployment.

E. Generalisation to Different Channel Conditions

The model is trained and tested using simulated data generated under the selected Rayleigh fading channel configuration. Its performance under other propagation environments has not been established. Practical wireless channels may differ in their fading distributions, antenna configurations, mobility conditions, interference levels, and temporal variations. A model that performs well under one simulated configuration may not maintain the same performance when these conditions change. The current results should therefore not be interpreted as evidence of generalisation across all Massive MIMO or 6G communication scenarios. Further testing is required to determine whether the learned channel-estimation behaviour remains reliable when the channel model or operating conditions differ from those used during training. Future studies should evaluate the model under multiple channel conditions, antenna configurations, and mobility scenarios. Testing with independent datasets and measured channel responses would provide stronger evidence of generalisation.

F. Experimental Reproducibility

Reproducing the reported results requires sufficient information about the simulation configuration, dataset generation, model architecture, and training procedure. The reported settings include the 8 × 8 MIMO configuration, 16-QAM modulation, Rayleigh fading, AWGN, 3,000 training samples, 500 testing samples, a batch size of 64, and 15 training epochs. However, additional implementation details are needed for a complete independent reproduction. These include the detailed attention architecture, the number of attention heads or blocks if applicable, the learning rate, the random seed, the definition of an individual sample, the pilot configuration, and the number of independent simulation trials. These details should be documented and verified against the implementation rather than inferred from the performance results. Future work should provide the complete model configuration, data-generation procedure, software environment, and evaluation protocol. Repeated experiments would also help establish the variability of the measured results.

G. Threats to Internal and External Validity

Internal validity concerns whether the observed differences can be attributed reliably to the methods being compared. Although the reported results show lower BER and SER for the attention-based model than for the FFNN baseline, the available evaluation does not establish the statistical significance of these differences. The number of independent runs and the variability across runs are not reported. External validity concerns whether the findings extend to conditions beyond the evaluated simulation. Since the experiments use a specific 8 × 8 Rayleigh fading configuration, the results may not transfer directly to other antenna sizes, modulation schemes, propagation environments, or practical communication systems. Accordingly, the results support a comparison within the evaluated experimental setup, but broader claims require additional validation. Repeated trials, uncertainty estimates, alternative channel models, and more diverse test conditions would strengthen the reliability and general applicability of the conclusions.

H. Validity of the Performance Measures

The study evaluates the methods using BER, SER, detection accuracy, and channel-estimation mean square error (MSE). These metrics provide complementary information, but they measure different aspects of system performance. Channel-estimation MSE measures the difference between the true channel and its estimate, whereas BER and SER measure errors in the recovered information. A reduction in channel-estimation MSE does not necessarily produce a proportional reduction in BER or SER because detection performance also depends on the channel configuration, noise, and detector. The reported results illustrate the importance of evaluating both estimation and detection performance. At 20 dB, the attention-based method performs better than the FFNN baseline, but MMSE remains the strongest method among the reported comparisons. The attention model also remains behind LS at higher SNR values. Future evaluations should retain these complementary metrics and add computational measurements and repeated-run analysis. This would provide a more comprehensive assessment of the proposed model's benefits and limitations. Overall, these limitations define the scope of the present findings. The experiments indicate that attention-based refinement can improve on the selected FFNN baseline, but additional training, architectural investigation, reproducibility measures, and validation under diverse conditions are necessary before the method can be considered a general replacement for conventional channel estimators.

VII. PRACTICAL CONSIDERATIONS

Before an attention-based channel estimator can be considered for practical deployment, its performance must be evaluated beyond BER and SER alone. Factors such as generalisation to real wireless channels, robustness across different SNR levels, training requirements, processing

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

latency, and integration with conventional receiver components must also be considered. These aspects are discussed below, but they were not experimentally evaluated in the present study.

A. Generalisation to Different Channel Conditions

The proposed model was trained and evaluated using a synthetic dataset generated under a Rayleigh fading channel model. Although this setup provides a controlled environment for comparing channel-estimation methods, it does not represent every condition encountered in an operational wireless network. Real propagation environments may include line-of-sight components, changing multipath conditions, user mobility, interference, and time-varying channel characteristics. A model trained under one set of channel conditions may not maintain the same performance when these conditions change. Its generalisation capability should therefore be evaluated using additional channel models and independently generated test data. Depending on the results, retraining or adaptation may be required for different deployment environments. Such robustness tests remain future work for the proposed model.

B. Performance Across the SNR Range

The simulation covers SNR values from 0 to 20 dB. The results show that the attention-based model improves upon the FFNN baseline throughout this range and performs better than LS estimation at 0 and 5 dB. However, its BER decreases only slightly at higher SNR values and remains substantially above the LS and MMSE results from 10 dB onward. This behaviour limits the model's suitability as a general replacement for conventional channel estimation. Its performance outside the evaluated SNR range is also unknown. Further experiments should investigate whether additional training data, changes to the network architecture, or explicit inclusion of SNR information can improve its response across different noise conditions. These are proposed investigations rather than verified solutions.

C. Training Data and Computational Cost

The model was trained using 3,000 synthetic samples for 15 epochs and evaluated using 500 test samples. This configuration establishes an initial experimental baseline, but it does not establish how the model would perform with larger datasets or more extensive training. In a practical system, training data would need to be obtained from reliable simulations, measurements, or a combination of both. Data generation, preprocessing, model training, and periodic retraining may introduce additional resource requirements. The effect of dataset size and training duration should therefore be examined systematically before drawing conclusions about the method's practical efficiency.

D. Processing Complexity and Latency

Low latency is an important requirement for many prospective 6G applications. Although an attention-based model may provide useful nonlinear refinement of an LS estimate, its suitability for real-time operation depends on the computational resources required for inference and on the time taken by the complete receiver pipeline. The present study did not measure inference latency, operation count, parameter count, memory usage, or hardware utilisation. Consequently, it cannot establish whether the proposed method satisfies a particular real-time processing requirement or is more computationally efficient than LS, MMSE, or the FFNN baseline. Future evaluations should measure these quantities under consistent hardware and software conditions. Both the attention-based refinement stage and the subsequent ZF detector should be included when assessing end-to-end receiver cost.

E. Integration with Conventional Receiver Methods

A learned estimator does not necessarily need to replace every conventional receiver component. It may instead operate as a refinement stage within a receiver that retains established signal-processing methods. In the proposed system, the attention network refines the initial LS channel estimate, and the refined estimate is supplied to a ZF detector. The results indicate that this arrangement improves upon the FFNN baseline, but it does not outperform MMSE estimation across the tested SNR range. The practical value of the refinement stage must therefore be judged by considering both its error-rate performance and its additional computational cost. One direction for future investigation is to use the refined channel estimate with an MMSE detector as well as with ZF. Such a comparison would help distinguish the contribution of channel-estimation quality from the effect of the selected detection algorithm. The outcome of this additional experiment cannot be inferred from the current results.

F. Interpretation and Reporting of Results

Practical claims must remain consistent with the scope of the available evidence. In this study, the attention-based model achieves lower BER and SER than the FFNN baseline at all evaluated SNR values, while its advantage over LS is limited to 0 and 5 dB. MMSE produces the lowest BER and SER throughout the tested range. These findings support the use of the model as a preliminary alternative to a plain FFNN under the simulated conditions;

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

they do not establish superiority over conventional estimators or readiness for deployment in a 6G network. Further evaluation of generalisation, processing cost, repeatability, and realistic channel conditions is necessary before broader practical conclusions can be drawn.

VIII. APPLICATIONS

The proposed attention-based channel-estimation method may be relevant to future wireless communication applications, including:

Table IX. Application Areas

Area	Potential Use
6G networks	Reliable wireless communication and signal detection
Smart cities	Connectivity for intelligent infrastructure
Autonomous vehicles	Communication under changing channel conditions
Industry 4.0	Wireless industrial monitoring and automation
Internet of Things	Reliable sensor connectivity
Remote healthcare	Data transmission for patient monitoring
Smart agriculture	Wireless agricultural sensor networks
Defence	Communication in demanding environments
Virtual and augmented reality	Reliable wireless data transmission

IX. CONCLUSION AND FUTURE WORK

A. Conclusion

This study investigated an attention-based deep learning model for pilot-assisted channel estimation and signal detection in an (8\times8) Massive MIMO system using 16-QAM modulation under Rayleigh fading and AWGN conditions. The model refines the initial least-squares (LS) channel estimate before signal detection. Experimental results demonstrated improved performance over the FFNN baseline across the evaluated SNR range of 0–20 dB. At 20 dB, the proposed model achieved a BER of 0.22525 and an SER of 0.39663, compared with 0.43219 and 0.68150, respectively, for the FFNN. However, the MMSE estimator consistently achieved better results, while LS outperformed the attention-based model at higher SNR values. The observed BER plateau also indicates that the current model has limitations in exploiting improvements in channel conditions. Therefore, although the approach shows potential as a learning-based alternative to a conventional neural network baseline, further optimization is necessary before it can compete effectively with established channel estimation techniques.

B. Future Work

Future research will focus on improving the model architecture, expanding the training dataset, and increasing the training duration to investigate the observed performance plateau. Providing SNR information as an additional model input may help the network adapt its channel refinement to different noise conditions. Further experiments will evaluate the refined channel estimate with an MMSE detector to distinguish channel-estimation performance from detection-related effects. The model will also be tested under different pilot configurations and channel conditions to assess its generalization capability. Repeated experiments with multiple random seeds will help evaluate the consistency of the results. Finally, computational complexity, parameter count, and processing latency will be measured and compared with LS, MMSE, and FFNN methods. These investigations will help determine the model's reliability, efficiency, and suitability for future Massive MIMO communication systems.

C. Assessment of Objectives

The attention-based model was successfully developed and evaluated. It improved detection performance compared with the FFNN baseline across the tested SNR range. However, it did not consistently outperform LS and MMSE estimators. Therefore, the objective of improving channel estimation and signal detection was partially achieved, with further optimization required.

APPENDIX A. ABBREVIATIONS

Short	Meaning
6G	Sixth Generation
AI	Artificial Intelligence
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
CSI	Channel State Information
DL	Deep Learning
FFNN	Feed-Forward Neural Network
LS	Least Squares
MIMO	Multiple Input Multiple Output
MMSE	Minimum Mean Square Error
MSE	Mean Square Error
QAM	Quadrature Amplitude Modulation
SER	Symbol Error Rate
SNR	Signal-to-Noise Ratio
ZF	Zero Forcing

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

APPENDIX B. NOTATION

Symbol	Meaning
N, M	Number of transmit and receive antennas (both 8)
$\mathbf{x}$, $\mathbf{y}$, $\mathbf{n}$	Transmitted vector, received vector, noise vector
$\mathbf{H}$	M×N channel matrix
$\mathbf{X}_p$, $\mathbf{Y}_p$	Pilot symbols and received pilots
$\hat{\mathbf{H}}_{LS}$	LS channel estimate
$\hat{\mathbf{H}}_{Att}$	Refined estimate from the attention model
f_{Att}	Trained attention network
$\mathbf{Q}$, $\mathbf{K}$, $\mathbf{V}$	Query, key and value projections
σ^2, E_s	Noise variance, mean symbol energy
B	Batch size
A	Set of all possible symbol vectors

REFERENCES

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," IEEE Transactions on Wireless Communications, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] F. Rusek et al., "Scaling up MIMO: Opportunities and challenges with very large arrays," IEEE Signal Processing Magazine, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[3] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," IEEE Transactions on Communications, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

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

[5] E. Björnson, J. Hoydis, and L. Sanguinetti, Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency. Boston, MA, USA: Now Publishers, 2017.

[6] T. L. Marzetta, "Massive MIMO: An introduction," Bell Labs Technical Journal, vol. 20, pp. 11–22, 2015.

[7] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," IEEE Journal on Selected Areas in Communications, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[8] R. R. Müller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," IEEE Journal of Selected Topics in Signal Processing, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[9] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," IEEE Transactions on Wireless Communications, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[10] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, Fundamentals of Massive MIMO. Cambridge, U.K.: Cambridge University Press, 2016.

[11] S. M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[12] S. M. Kay, Fundamentals of Statistical Signal Processing: Detection Theory. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.

[13] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," IEEE Transactions on Broadcasting, vol. 48, no. 3, pp. 223–229, Sep. 2002.

[14] O. Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson, and P. O. Börjesson, "OFDM channel estimation by singular value decomposition," IEEE Transactions on Communications, vol. 46, no. 7, pp. 931–939, Jul. 1998.

[15] M. Morelli and U. Mengali, "A comparison of pilot-aided channel estimation methods for OFDM systems," IEEE Transactions on Signal Processing, vol. 49, no. 12, pp. 3065–3073, Dec. 2001.

[16] M. K. Ozdemir and H. Arslan, "Channel estimation for wireless OFDM systems," IEEE Communications Surveys & Tutorials, vol. 9, no. 2, pp. 18–48, 2007.

[17] A. Goldsmith, Wireless Communications. Cambridge, U.K.: Cambridge University Press, 2005.

[18] D. Tse and P. Viswanath, Fundamentals of Wireless Communication. Cambridge, U.K.: Cambridge University Press, 2005.

[19] A. F. Molisch, Wireless Communications, 2nd ed. Chichester, U.K.: Wiley, 2011.

[20] A. Paulraj, R. Nabar, and D. Gore, Introduction to Space-Time Wireless Communications. Cambridge, U.K.: Cambridge University Press, 2003.

[21] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Model-driven deep learning for MIMO detection," IEEE Transactions on Signal Processing, vol. 68, pp. 1702–1715, 2020.

[22] N. Samuel, T. Diskin, and A. Wiesel, "Deep MIMO detection," in Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2017, pp. 1–5.

[23] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," IEEE Wireless Communications Letters, vol. 7, no. 1, pp. 114–117, Feb. 2018.

[24] H. Ye, G. Y. Li, and B.-H. Juang, "Deep learning-based end-to-end wireless communication systems with conditional GANs as unknown channels," IEEE Transactions on Wireless Communications, vol. 19, no. 5, pp. 3133–3143, May 2020.

[25] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," IEEE Transactions on Cognitive Communications and Networking, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[26] N. Farsad and A. Goldsmith, "Neural network detection of data sequences in communication systems," IEEE Transactions on Signal Processing, vol. 66, no. 21, pp. 5663–5678, Nov. 2018.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

[27] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," IEEE Journal of Selected Topics in Signal Processing, vol. 12, no. 1, pp. 168–179, Feb. 2018.

[28] C.-J. Chun, J.-M. Kang, and I.-M. Kim, "Deep learning-based channel estimation for massive MIMO systems," IEEE Wireless Communications Letters, vol. 8, no. 4, pp. 1228–1231, Aug. 2019.

[29] M. Soltani, V. Pourahmadi, A. Mirzaei, and H. Sheikhzadeh, "Deep learning-based channel estimation," IEEE Communications Letters, vol. 23, no. 4, pp. 652–655, Apr. 2019.

[30] P. Dong, H. Zhang, G. Y. Li, I. Gaspar, and N. Naderializadeh, "Deep CNN-based channel estimation for mmWave massive MIMO systems," IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 5, pp. 989–1000, Sep. 2019.

[31] E. Balevi, A. Doshi, and J. G. Andrews, "Massive MIMO channel estimation with an untrained deep neural network," IEEE Transactions on Wireless Communications, vol. 19, no. 3, pp. 2079–2090, Mar. 2020.

[32] J. Gao, M. Hu, C. Zhong, G. Y. Li, and Z. Zhang, "An attention-aided deep learning framework for massive MIMO channel estimation," IEEE Transactions on Wireless Communications, 2021.

[33] H. Hirose, T. Ohtsuki, and G. Gui, "Deep learning-based channel estimation for massive MIMO systems with pilot contamination," IEEE Open Journal of the Vehicular Technology Society, 2021.

[34] L. Ge et al., "Deep neural network based channel estimation for massive MIMO-OFDM systems with imperfect channel state information," IEEE Systems Journal, vol. 16, no. 3, pp. 4675–4685, 2022.

[35] J. Guo, T. Chen, S. Jin, G. Y. Li, X. Wang, and X. Hou, "Deep learning for joint channel estimation and feedback in massive MIMO systems," Digital Communications and Networks, vol. 10, no. 1, pp. 83–93, 2024.

[36] A. Vaswani et al., "Attention is all you need," in Advances in Neural Information Processing Systems, vol. 30, 2017, pp. 5998–6008.

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.

[38] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4700–4708.

[39] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7132–7141.

[40] X. Wang et al., "Non-local neural networks," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7794–7803.

[41] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. International Conference on Learning Representations (ICLR), 2015.

[42] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in Proc. International Conference on Artificial Intelligence and Statistics (AISTATS), 2011, pp. 315–323.

[43] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, May 2015.

[44] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[45] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[46] A. Vaswani et al., "Attention is all you need," in Advances in Neural Information Processing Systems, vol. 30, 2017.

[47] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," Foundations and Trends in Signal Processing, vol. 11, nos. 3–4, pp. 154–655, 2017.

[48] T. S. Rappaport et al., "Millimeter wave mobile communications for 5G cellular: It will work!" IEEE Access, vol. 1, pp. 335–349, 2013.

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

[50] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions," IEEE Open Journal of the Communications Society, vol. 1, pp. 957–975, 2020.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)