Engineering Systems and Intelligent Technologies
ESIT
ISSN: 3071-253X/
Journal Homepage
https://pub.scientificirg.com/index.php/ESIT

Robust LSTM-Based Channel State Estimation for Next-Generation Wireless Communication Systems

M. S. Yasseen [a,1] **, and Hana Mujlid** [b]

[a] *Department of Electrical Engineering, Faculty of Engineering, Al-Azhar University, Qena 83513, Egypt; E-mail: myasseen@azhar.edu.eg*

[b] *Department of Computer Engineering, Taif University, 21944, Taif, Saudi Arabia; Email: hmujlid@tu.edu.sa*

ABSTRACT

Long Short-Term Memory (LSTM) networks conventionally employ the hyperbolic tangent (*tanh*) and sigmoid functions for state and gate activations, respectively. Although numerous alternative activation functions have been introduced for deep neural networks, their potential to improve LSTM performance in wireless communication tasks remains insufficiently explored. This paper investigates the replacement of the *tanh* state activation function in LSTM architectures with three alternatives: the Bi-tanh1, the Elliott, and the Gaussian Error Linear Unit (GELU). The resulting modified LSTM networks are evaluated as channel state estimators (CSEs) for 5G Orthogonal Frequency-Division Multiplexing (OFDM) systems operating over Rayleigh fading channels. Simulation results demonstrate that all proposed activation-function-based LSTM CSEs substantially outperform conventional Least Squares (LS) and Minimum Mean Square Error (MMSE) estimators, particularly under low pilot-density conditions. Among the evaluated configurations, the GELU-based LSTM CSE achieves the lowest symbol error rate (SER) across all tested signal-to-noise ratios (SNRs), establishing GELU as a strong candidate for robust channel estimation in next-generation wireless receivers.

PAPER INFORMATION

HISTORY

Received: 28 March 2026
Revised: 25 May 2026
Accepted: 11 June 2026
Online: 29 June 2026

MSC

68T07; 68T09; 94A12; 68M10; 94A08

KEYWORDS

LSTM;
Channel State Estimation;
5G Communication Systems;
Gaussian Error Linear Unit;
OFDM.

1. INTRODUCTION

The Long Short-Term Memory (LSTM) recurrent neural network (RNN), originally proposed by Hochreiter and Schmidhuber [1], has become a cornerstone of sequence-learning research owing to its ability to capture long-range temporal dependencies through gated memory cells. Each LSTM block contains one or more self-connected memory cells together with input, forget, and output gates. The gating mechanism enables the network to selectively store and retrieve information over arbitrarily long time intervals, alleviating the vanishing and exploding gradient problems that afflict standard RNNs.

The generalization capability of a deep learning network depends on several interacting factors, including network topology, regularization strategies, optimization algorithms, and the choice of activation function at each computational node [2, 3]. In practice, however, research attention has largely concentrated on architectural design and training procedures, leaving activation function selection comparatively unexplored [4]. Activation functions define the nonlinear decision

[1]Corresponding author at Department of Electrical Engineering, Faculty of Engineering, Al-Azhar University, Qena 83513, Egypt; E-mail: myasseen@azhar.edu.eg

boundary at each node, govern gradient propagation during backpropagation, and directly influence convergence speed and generalization accuracy. Developing better-suited activation functions can therefore accelerate training convergence and improve robustness against gradient pathologies [5, 6].

In the domain of fifth-generation (5G) wireless communications, LSTM-based deep learning has gained considerable traction [7, 8, 9]. Channel state estimation (CSE) is a fundamental task in OFDM receivers: accurate knowledge of the channel frequency response is required to equalize the distortion introduced by multipath propagation. Conventional pilot-assisted estimators such as Least Squares (LS) and Minimum Mean Square Error (MMSE) suffer from well-known limitations under low pilot-density and low signal-to-noise ratio (SNR) conditions. Specifically, LS estimation amplifies noise, while MMSE requires precise knowledge of second-order channel statistics that may not be available at the receiver [7].

Deep learning approaches can implicitly learn channel statistics from data, offering an appealing alternative. LSTM networks are particularly attractive because channel coefficients in time-varying multipath environments exhibit temporal correlation that recurrent architectures can exploit naturally [10, 11]. Prior work has demonstrated that replacing or modifying the default activation functions in LSTM can yield meaningful performance improvements in classification and regression tasks [12, 5]. Yet, this direction has not been systematically applied to 5G channel state estimation using pilot-based OFDM systems.

The present study addresses this gap by evaluating three non-standard state activation functions as drop-in replacements for the canonical *tanh* inside standard LSTM cells, while preserving the original sigmoid gate activations. The resulting LSTM-based CSEs are trained offline and deployed online to simultaneously perform channel estimation and symbol detection in a 16-QAM OFDM system over Rayleigh fading channels. Performance is benchmarked against LS and MMSE baselines under three pilot-density settings: high (64 pilots), medium (16 pilots), and extremely sparse (2 pilots).

To the best of the authors' knowledge, this is the first study to apply advanced activation-function-based LSTM networks specifically developed in [12] to the construction of robust 5G-OFDM receivers.

The principal contributions of the paper are as follows.

1) Three alternative state activation functions (Bi-tanh1, Elliott, and GELU) are investigated as replacements for *tanh* in LSTM-based CSEs.

2) The proposed LSTM CSEs implicitly perform channel estimation and symbol detection simultaneously, without requiring an explicit channel estimation stage.

3) Comprehensive SER comparisons under varying pilot densities demonstrate the superiority of activation-function-based LSTM CSEs over conventional LS and MMSE estimators.

4) Numerical results identify GELU as the most effective state activation function for LSTM-based CSE under high and medium pilot-density conditions.

The remainder of this paper is organized as follows. Section 2 surveys related work. Section 3 describes the LSTM architecture and the investigated activation functions. Section 4 presents the proposed methodology. Section 5 reports and analyzes the simulation results. Section 6 concludes the paper and outlines directions for future research.

2. RELATED WORK

2.1 Deep Learning for OFDM Channel Estimation

Pilot-assisted channel estimation in OFDM systems has been an active research area for several decades. Classical approaches rely on interpolation between pilot sub-carriers using LS or MMSE criteria. LS estimation offers low complexity but amplifies noise in frequency-selective fading channels. MMSE estimation achieves optimal linear performance but requires statistical knowledge of the channel and the noise variance, which may not be available in practice [7].

Deep neural networks have been proposed as data-driven alternatives to these classical estimators. Early contributions applied fully connected networks to learn the mapping from received pilot observations to estimated channel coefficients. Li et al. [13] proposed a deep learning framework for joint channel estimation and signal detection in OFDM systems, reporting SER improvements over LS and MMSE baselines. Sebin and Jose [11] introduced an LSTM with a projected layer to reduce model complexity while maintaining competitive estimation accuracy. Atallah [10] further proposed a peephole LSTM variant that outperformed standard LSTM and Gated Recurrent Unit (GRU) estimators under sparse pilot settings. Shankar et al. [14] evaluated a bidirectional LSTM (Bi-LSTM) for massive MIMO-OFDM over TDL-C channel models, demonstrating the value of bidirectional temporal context for channel tracking.

The above contributions confirm the capacity of recurrent deep learning architectures to outperform classical estimators, particularly in low-SNR and sparse-pilot regimes. However, none of these works systematically investigate the role of the state activation function within the LSTM memory cell.

2.2 Activation Functions in Deep Learning

The selection of appropriate activation functions has a decisive influence on the convergence and generalization of deep networks. Dubey et al. [3] provided a comprehensive survey and benchmark of over 400 activation functions, noting that no single function is universally optimal and that the best choice is task- and architecture-dependent.

For general deep networks, the Gaussian Error Linear Unit (GELU), introduced by Hendrycks and Gimpel [6], has emerged as a high-performing alternative to ReLU in natural language processing and computer vision tasks. GELU weights input values by their magnitude under a Gaussian prior, producing a smooth, non-monotone nonlinearity that avoids the dead-neuron phenomenon of ReLU while retaining good gradient flow. The Swish function [15] and related variants have similarly shown improvements over ReLU in deep architectures, motivating broader exploration of smooth activation functions.

In the specific context of recurrent networks, Vijayaprabakaran and Sathiyamurthy [5] applied a differential-evolution-based search to identify improved state activation functions for LSTM networks. Ali et al. [12] conducted the most comprehensive study to date, evaluating 26 candidate state activation functions across classification tasks and demonstrating that several alternatives, including Bi-tanh1, Elliott, and GELU, consistently outperform the default *tanh*. The present paper extends this line of investigation to the channel estimation task in 5G-OFDM systems.

2.3 LSTM for Wireless Communications

Beyond the applications mentioned in the introduction, LSTM architectures have been applied to channel prediction for 5G Massive MIMO systems [16], to the review of deep learning techniques across coding, multiple access, and resource allocation [8], and to channel equalization in Single Carrier Frequency Division Multiple Access (SC-FDMA) systems [17]. Hassan et al. [2, 9] proposed LSTM-based estimators for OFDM systems and reported that deep learning estimators can approach and sometimes exceed MMSE performance with fewer pilots. The current paper builds on these findings by demonstrating that activation function selection provides an additional degree of freedom for improving LSTM-based CSE performance.

3. LSTM ARCHITECTURE AND ACTIVATION FUNCTIONS

3.1 Standard LSTM Cell

Figure 1 presents the architecture of a standard LSTM memory cell. The cell is governed by the following six equations:

$$i_t = \sigma(W_i\, x_t + R_i\, h_{t-1} + b_i)\,, \tag{1}$$

$$f_t = \sigma\big(W_f\, x_t + R_f\, h_{t-1} + b_f\big)\,, \tag{2}$$

$$o_t = \sigma(W_o\, x_t + R_o\, h_{t-1} + b_o)\,, \tag{3}$$

$$g_t = \phi(W_x\, x_t + R_x\, h_{t-1} + b_x)\,, \tag{4}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t, \tag{5}$$

$$h_t = o_t \odot \phi(c_t), \tag{6}$$

where x_t is the input vector at time step t; i_t, f_t, and o_t are the input, forget, and output gate vectors, respectively; g_t is the block input (cell input activation); c_t is the cell state; h_t is the block output (hidden state); and h_{t-1} is the hidden state from the previous time step. The matrices W_* and R_* denote the input and recurrent weight matrices, and b_* are the bias vectors for each gate. The function $\sigma(\cdot)$ is the logistic sigmoid used for gating, and $\phi(\cdot)$ is the state activation function (SAF), which defaults to *tanh* in the standard LSTM formulation. The operator $\odot$ denotes element-wise (Hadamard) multiplication.

As seen in **Equation 4** and **Equation 6**, $\phi(\cdot)$ appears twice per time step: once in the computation of the block input g_t, and once in the output projection h_t. It is this function that is replaced in the experiments reported below.

In **Figure 1**, the gates i_t, f_t, and o_t use sigmoid activations $\sigma(\cdot)$, while the state activation function $\phi(\cdot)$ is applied to the block input g_t and the cell state output c_t before the hidden state h_t is produced.

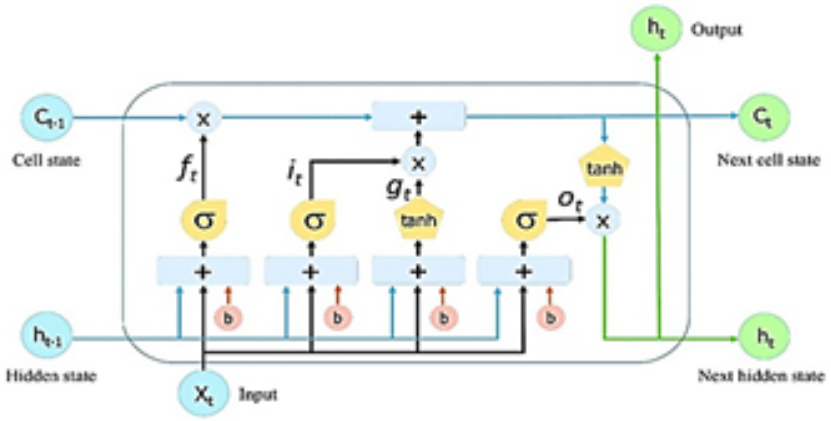

Figure 1: Detailed architecture of a standard LSTM memory cell

3.2 Candidate State Activation Functions

Table 1 lists the five activation functions examined in this study. The canonical *tanh* serves as the baseline. The three candidates proposed as replacements are Bi-tanh1, Elliott, and GELU. Softsign is included as an additional comparator because of its structural similarity to Elliott.

Table 1: Candidate state activation functions and their analytical derivatives

Function	**Expression** $\phi(x)$	**Derivative** $\phi'(x)$
tanh	$\dfrac{e^x - e^{-x}}{e^x + e^{-x}}$	$1 - \tanh^2(x)$
Bi-tanh1 [18]	$\dfrac{1}{2}\left[\tanh\left(\dfrac{x}{2}\right) + \tanh\left(\dfrac{x+1}{2}\right)\right] + \dfrac{1}{2}$	$\dfrac{\operatorname{sech}^2\left(\dfrac{x+1}{2}\right) + \operatorname{sech}^2\left(\dfrac{x}{2}\right)}{4}$
Softsign [12]	$\dfrac{x}{1+\lvert x\rvert} + \dfrac{1}{2}$	$\dfrac{1}{(1+\lvert x\rvert)^2}$
Elliott [12]	$\dfrac{0.5x}{1+\lvert x\rvert} + \dfrac{1}{2}$	$\dfrac{0.5}{(1+\lvert x\rvert)^2}$
GELU [6]	$\dfrac{x}{2}\left[1 + \operatorname{erf}\left(\dfrac{x}{\sqrt{2}}\right)\right]$	$\dfrac{1}{2}\left[1 + \operatorname{erf}\left(\dfrac{x}{\sqrt{2}}\right)\right] + \dfrac{x\, e^{-x^2/2}}{\sqrt{2\pi}}$

Brief descriptions of each function are given below.

tanh:

The hyperbolic tangent is the default SAF in standard LSTM implementations. Its output is bounded to $(-1, 1)$, and its derivative reaches a maximum of 1 at $x = 0$, decaying toward zero for large $|x|$, which can contribute to the vanishing gradient problem in deep or long-horizon recurrent networks.

Bi-tanh1:

Introduced by Sodhi and Chandra [18], the Bi-tanh1 function is constructed as the sum of two shifted *tanh* functions. Its derivative exhibits two local maxima, which broadens the activation range and can improve gradient flow during backpropagation through time. The output range of Bi-tanh1 is $(0, 1)$, which differs from the zero-centered $(-1, 1)$ range of *tanh*.

Elliott:

The Elliott function [12] is a rational approximation to the sigmoid family. Its derivative is computationally inexpensive because it avoids transcendental operations and depends only on $|x|$. The function range is $(0, 1)$.

Softsign:

Softsign is structurally related to Elliott, differing only in the numerator coefficient. Its derivative decays more slowly than that of *tanh*, which may help sustain gradient magnitudes over long sequences.

GELU:

GELU was introduced by Hendrycks and Gimpel [6] and is defined as $\phi(x) = x\,\Phi(x)$, where $\Phi(x)$ is the standard Gaussian cumulative distribution function. An exact expression in terms of the error function is listed in **Table 1**. GELU is smooth, non-monotone for negative inputs, and weights each input by the probability that a Gaussian random variable exceeds it. This stochastic interpretation endows GELU with desirable gradient properties and has led to its adoption as the default activation function in large-scale language models and vision transformers.

4. PROPOSED METHODOLOGY

4.1 System Model

This study replaces the conventional tanh state activation function (listed in **Table 1**) to investigate its impact on channel state estimation effectiveness. The sigmoid gate activation functions remain unchanged. The novel LSTM-based channel state estimators (CSEs) are trained using the Adam optimization algorithm. Training continues until the LSTM-based CSEs achieve a low and approximately constant symbol error rate (SER) on the training data. All examined DNN configurations employ a categorical cross-entropy loss function at the final classification layer.

4.2 LSTM Configuration and Activation Function Substitution

The state activation function $\phi(\cdot)$ in **Equation** (4) and **Equation** (6) is replaced by each candidate listed in **Table 1**, one at a time. The sigmoid gate activations in **Equations 1–3** are left unchanged to isolate the contribution of the SAF. Each modified LSTM configuration is denoted LSTM_{ϕ}, where ϕ identifies the SAF. For example, $\text{LSTM}_{\text{GELU}}$ denotes the LSTM cell with GELU as the state activation function.

4.3 Offline Training Procedure

Each LSTM_{ϕ} is trained offline on a simulated dataset before deployment. The training procedure follows the specifications established in [7]. The network receives the complex-valued received OFDM signal as a sequence of real and imaginary parts and outputs a probability vector over the 16-QAM symbol alphabet. A categorical cross-entropy loss function is minimized using the Adam optimizer [19] with a fixed learning rate of 0.001 and a mini-batch size of 128. Training proceeds for 100 epochs or until the training SER stabilizes at a low value. The complete set of simulation parameters is listed in **Table 2**.

5. SIMULATION RESULTS

Comprehensive experiments are conducted to compare the SER performance of five LSTM_{ϕ} configurations (using *tanh*, Bi-tanh1, Softsign, Elliott, and GELU as the SAF) against conventional LS and MMSE estimators. Three pilot-density scenarios are evaluated: 64 pilots (full-density, Section 5.1), 16 pilots (quarter-density, Section 5.2), and 2 pilots (extreme sparsity, Section 5.3). In all figures, E_s/N_0 (dB) on the horizontal axis represents the received signal energy per symbol to noise power spectral density ratio.

Table 2: OFDM system parameters and LSTM training configuration

Parameter	Value
FFT size (N)	64
Cyclic prefix length	16
Modulation scheme	16-QAM
Channel model	Rayleigh fading, $L = 5$ paths
Training SNR range	0–20 dB
Optimizer	Adam [19]
Learning rate	0.001
Mini-batch size	128
Training epochs	100
Loss function	Categorical cross-entropy

5.1 Performance with 64 Pilots

Figure 2 presents the SER curves for the full pilot-density setting of $N_p = 64$. At this density, conventional estimators have access to a complete set of pilot observations, providing the most favorable conditions for LS and MMSE.

Despite this advantage, all five LSTM$_\phi$ configurations substantially outperform the LS estimator across the entire SNR range, and they match or exceed MMSE performance at most operating points. Among the five configurations, LSTM$_{\text{GELU}}$ achieves the best SER, followed by LSTM$_{\text{Elliott}}$ and LSTM$_{\text{Bi-tanh1}}$. LSTM$_{\text{Softsign}}$ and the baseline LSTM$_{\text{tanh}}$ perform somewhat worse but still clearly surpass the conventional estimators. This ordering suggests that both the smoothness of GELU and the broad derivative range of Elliott are beneficial for the channel estimation task when pilot coverage is dense.

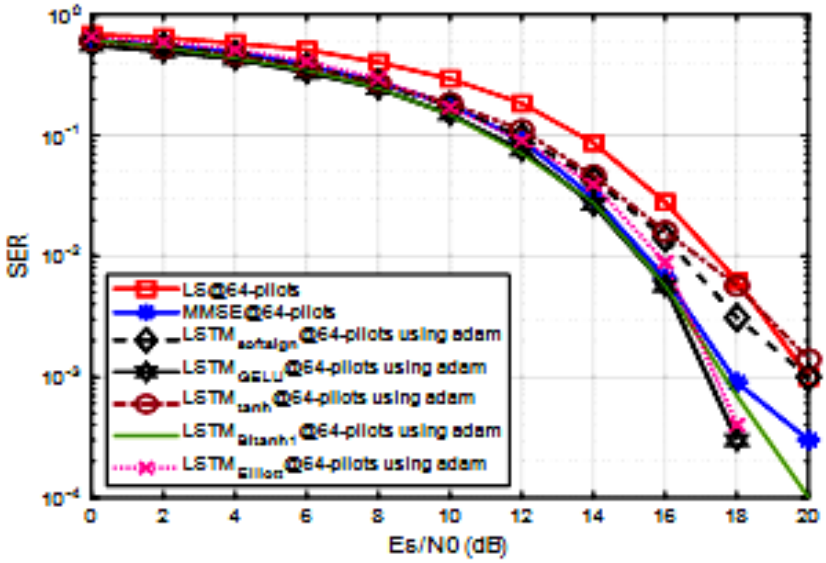

Figure 2: SER comparison of proposed SAF-based LSTM CSEs against conventional LS and MMSE estimators using $N_p = 64$ pilots

5.2 Performance with 16 Pilots

Figure 3 shows results for the quarter-density setting, $N_p = 16$. Reducing pilot density degrades the performance of conventional estimators markedly, particularly at high SNR where interpolation errors limit the error floor. The LSTM configurations are considerably more robust; all five maintain the performance hierarchy observed in the 64-pilot case.

At $E_s/N_0 = 18$ dB, quantitative SER values are as follows: LSTM$_{\text{GELU}}$: 9.0×10^{-4}; LSTM$_{\text{Elliott}}$: 6.0×10^{-4}; LSTM$_{\text{Bi-tanh1}}$: 7.0×10^{-4}; LSTM$_{\text{tanh}}$: 1.3×10^{-3}; MMSE: 1.94×10^{-2}; LS: 3.26×10^{-2}. The proposed LSTM CSEs thus outperform the MMSE estimator by approximately one order of magnitude in SER at 18 dB, and by a factor of approximately 35 over LS. The relatively strong performance of LSTM$_{\text{Elliott}}$ in this regime can be attributed to the stable gradient magnitude provided by its rational derivative form, which may improve convergence under data-limited training

conditions.

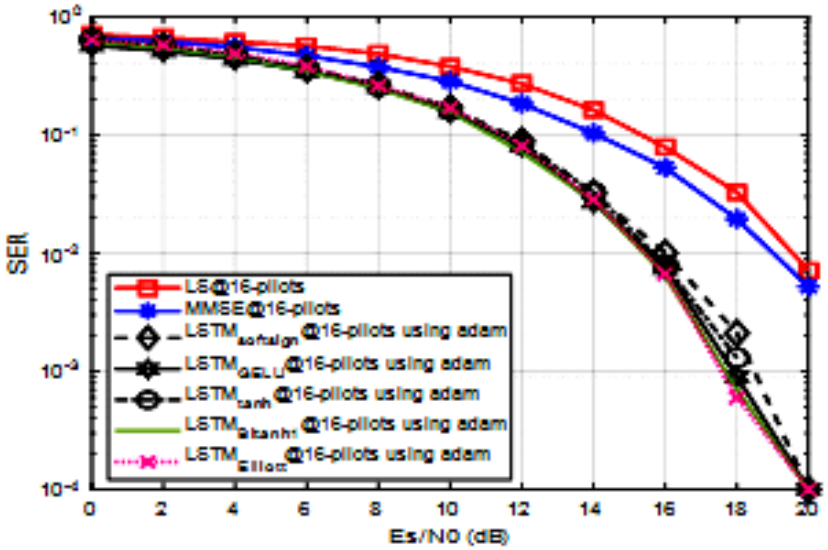

Figure 3: SER comparison of proposed SAF-based LSTM CSEs against conventional LS and MMSE estimators using $N_p = 16$ pilots

5.3 Performance with 2 Pilots

Figure 4 illustrates the most challenging scenario, $N_p = 2$. Under this extreme pilot-sparsity condition, both the LS and MMSE estimators effectively fail, returning SER values near 10^{-1} even at high SNR, because two pilot observations are insufficient for accurate 64-sub-carrier channel interpolation. The $\text{LSTM}_{\text{Elliott}}$ configuration also loses effectiveness under these conditions, with SER values tracking closely with those of the conventional estimators, suggesting that the Elliott function is less suitable when training signal coverage is extremely limited.

The remaining four LSTM configurations demonstrate competitive SER performance in the low-SNR region (0–6 dB). At $E_s/N_0 = 18$ dB, $\text{LSTM}_{\text{tanh}}$ achieves the lowest SER of 9.0×10^{-4}, followed by $\text{LSTM}_{\text{Bi-tanh1}}$ at 1.3×10^{-3}, while $\text{LSTM}_{\text{GELU}}$ and $\text{LSTM}_{\text{Softsign}}$ are competitive at 16 dB. The Bi-tanh1 function achieves notably strong performance at 16 dB, outperforming GELU in this intermediate SNR range.

The reversal of the performance ordering compared to the denser-pilot scenarios is noteworthy. Under extreme data scarcity, the zero-centered output range of *tanh* and the broad derivative profile of Bi-tanh1 appear better suited to learning from minimal pilot information. This finding motivates further investigation of activation function selection as a function of the training data regime.

5.4 Comparative Analysis

Table 3 summarizes the relative performance ranking of the five LSTM configurations across the three pilot-density scenarios. $\text{LSTM}_{\text{GELU}}$ is the top performer in both the 64-pilot and 16-pilot settings, and remains competitive in the 2-pilot setting. $\text{LSTM}_{\text{Elliott}}$ ranks second in the medium-density scenario but degrades under extreme sparsity. $\text{LSTM}_{\text{Bi-tanh1}}$ exhibits consistent mid-to-upper ranking across all three settings. These observations suggest that the choice of activation function should take into account the available pilot density in the target deployment scenario.

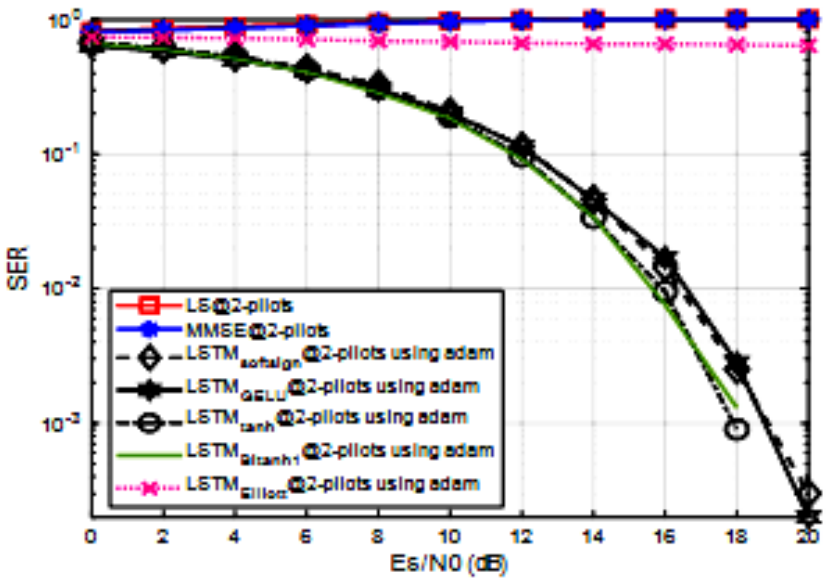

Figure 4: SER comparison of proposed SAF-based LSTM CSEs against conventional LS and MMSE estimators using $N_p = 2$ pilots

Table 3: Performance ranking of LSTM$_\phi$ configurations by SER at $E_s/N_0 = 18$ dB across three pilot-density settings. Rank 1 denotes the lowest SER

Configuration	64 Pilots	16 Pilots	2 Pilots
$\text{LSTM}_{\text{GELU}}$	1	2	3
$\text{LSTM}_{\text{Elliott}}$	2	1	5
$\text{LSTM}_{\text{Bi-tanh1}}$	3	3	2
$\text{LSTM}_{\text{Softsign}}$	4	4	4
$\text{LSTM}_{\text{tanh}}$	5	5	1

6. CONCLUSION

This paper demonstrated that the selection of the state activation function inside LSTM memory cells has a significant impact on channel state estimation performance in 5G-OFDM systems. Replacing the default *tanh* activation with Bi-tanh1, Elliott, or GELU yields measurable SER improvements under high and medium pilot-density conditions. Among the evaluated configurations, $\text{LSTM}_{\text{GELU}}$ achieves the best overall SER across E_s/N_0 values from 0 to 20 dB when pilot density is sufficient (64 or 16 pilots), outperforming conventional LS and MMSE estimators by roughly one order of magnitude. Under extreme pilot sparsity (2 pilots), the conventional *tanh* and Bi-tanh1 functions prove more resilient, highlighting the importance of matching activation function choice to the deployment pilot configuration.

The proposed approach offers a practical advantage: no modification to the LSTM network topology, loss function, or training algorithm is required; only the state activation function is changed. This low-overhead intervention can readily be integrated into existing LSTM-based receiver designs.

Several directions merit follow-up study. Gate activation functions and recurrent weight initialization strategies within the LSTM cell have not been explored and may offer complementary gains. Extension to higher-order modulation schemes (64-QAM and 256-QAM), massive MIMO-OFDM channels, and channel models beyond Rayleigh fading (e.g., Rician, doubly dispersive vehicular channels) would broaden the applicability of the findings. Adaptive or data-driven selection of the activation function based on pilot density or estimated channel statistics is also a promising avenue. Finally, the suitability of GELU and Bi-tanh1 as state activation functions warrants investigation in other sequential deep learning applications, including time-series forecasting, speech processing, and online handwriting recognition.

AUTHOR CONTRIBUTION STATEMENT

All authors contributed equally to the study conception and design. Material preparation, data collection, and analysis were performed by the authors. The first draft of the manuscript was written by the authors, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Ethics declaration: not applicable. This study did not involve human participants or animals. Therefore, ethical approval and consent to participate are not applicable.

CONSENT FOR PUBLICATION

Consent to Publish declaration: not applicable.

DATA AVAILABILITY

The dataset supporting the findings of this study was generated synthetically through simulations.

ACKNOWLEDGMENTS

The author gratefully acknowledges the intellectual contribution of the broader interdisciplinary literature reviewed in this article. No external funding was received for this study.

FUNDING

No Funding.

DISCLOSURE STATEMENT

The author declares no conflict of interest. The article is based on a structured review of published literature and did not involve human participants, personal data collection, or experimental intervention.

REFERENCES

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

[2] H. A. Hassan, M. A. Mohamed, M. H. Essai, H. Esmaiel, A. S. Mubarak, and O. A. Omer, "Effective deep learning-based channel state estimation and signal detection for ofdm wireless systems," *Journal of Electrical Engineering*, vol. 74, no. 3, pp. 167–176, 2023.

[3] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, "Activation functions in deep learning: A comprehensive survey and benchmark," *Neurocomputing*, vol. 503, pp. 92–108, 2022.

[4] M. Basirat and P. M. Roth, "Learning task-specific activation functions using genetic programming.," in *VISIGRAPP (5: VISAPP)*, pp. 533–540, 2019.

[5] K. Vijayaprabakaran and K. Sathiyamurthy, "Towards activation function search for long short-term model network: A differential evolution based approach," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 6, pp. 2637–2650, 2022.

[6] D. Hendrycks and K. Gimpel, "Gaussian error linear units (gelus)," *arXiv preprint arXiv:1606.08415*, 2016.

[7] M. H. E. Ali and I. B. Taha, "Channel state information estimation for 5g wireless communication systems: recurrent neural networks approach," *PeerJ Computer Science*, vol. 7, p. e682, 2021.

[8] A. Ly and Y.-D. Yao, "A review of deep learning in 5g research: Channel coding, massive mimo, multiple access, resource allocation, and network security," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 396–408, 2021.

[9] H. A. Hassan, M. A. Mohamed, M. N. Shaaban, M. H. E. Ali, and O. A. Omer, "An efficient deep neural network channel state estimator for ofdm wireless systems," *Wireless Networks*, vol. 30, no. 3, pp. 1441–1451, 2024.

[10] M. H. Essai Ali, A. R. Abdellah, H. A. Atallah, G. S. Ahmed, A. Muthanna, and A. Koucheryavy, "Deep learning peephole lstm neural network-based channel state estimators for ofdm 5g and beyond networks," *Mathematics*, vol. 11, no. 15, p. 3386, 2023.

[11] S. J. Olickal and R. Jose, "Lstm projected layer neural network-based signal estimation and channel state estimator for ofdm wireless communication systems.," *AIMS Electronics & Electrical Engineering*, vol. 7, no. 2, 2023.

[12] M. H. E. Ali, A. B. Abdel-Raman, and E. A. Badry, "Developing novel activation functions based deep learning lstm for classification," *IEEE Access*, vol. 10, pp. 97259–97275, 2022.

[13] J. Li, Z. Zhang, Y. Wang, B. He, W. Zheng, and M. Li, "Deep learning-assisted ofdm channel estimation and signal detection technology," *IEEE Communications Letters*, vol. 27, no. 5, pp. 1347–1351, 2023.

[14] R. Shankar, "Bi-directional lstm based channel estimation in 5g massive mimo ofdm systems over tdl-c model with rayleigh fading distribution," *International Journal of Communication Systems*, vol. 36, no. 16, p. e5585, 2023.

[15] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," *arXiv preprint arXiv:1710.05941*, 2017.

[16] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, "Channel state information prediction for 5g wireless communications: A deep learning approach," *IEEE transactions on network science and engineering*, vol. 7, no. 1, pp. 227–236, 2018.

[17] M. A. Mohamed, H. A. Hassan, M. H. Essai, H. Esmaiel, A. S. Mubarak, and O. A. Omer, "Modified gate activation functions of bi-lstm-based sc-fdma channel equalization," *Journal of Electrical Engineering*, vol. 74, no. 4, pp. 256–266, 2023.

[18] S. S. Sodhi and P. Chandra, "Bi-modal derivative activation function for sigmoidal feedforward networks," *Neurocomputing*, vol. 143, pp. 182–196, 2014.

[19] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.