Feedback Deep Neural Networks for Enhanced Multi-User Detection in MIMO-NOMA Systems

Author 1: **Dr. Dilip Kumar Sharma**, Faculty, Ujjain Engineering College, India. Email: drdilipsharma72@gmail.com

Author 2**: Nidhi Sisodiya**, PG student, Ujjain Engineering College, India. Email: nidhi.sis1998@gmail.com

Abstract—non-orthogonal multiple access (NOMA) integrated with multiple-input multiple-output (MIMO) technology is pivotal for enhancing spectral efficiency in fifth-generation (5G) and sixth-generation (6G) wireless networks. Conventional successive interference cancellation (SIC) receivers, however, are plagued by issues such as near-far effects and error propagation, resulting in high bit error rates (BER), particularly for users with minimal signal-to-noise ratio (SNR) disparities. This paper introduces a feedback deep neural network (FDNN) receiver that employs deep learning for nonlinear signal detection and incorporates iter-

Through extensive simulations, the FDNN demonstrates a substantial BER reduction for the weaker user, from 9.8×10^{-2} to 2.4×10^{-2} at 20 dB SNR, surpassing zero-forcing (ZF)-SIC baselines. The study extends evaluations with ablation studies on network architectures, batch sizes, increased user counts (up to K=3), modulation schemes (QPSK and 16-QAM), and performance under imperfect channel state information (CSI). Additional analyses include computational complexity, latency metrics, and robustness in dynamic channel environments. The FDNN's scalable and interference-resilient design holds signif- icant promise for next-generation wireless systems, addressing key challenges in real-time deployment.

Index Terms—MIMO-NOMA, deep neural network, feed-back receiver, successive interference cancellation, bit error rate, 5G/6G, imperfect CSI, deep learning.

I. Introduction II. Background and Motivation

The evolution of wireless communication systems to- wards 5G and beyond necessitates innovative multiple access techniques to accommodate exponential growth in connected devices and data traffic. Non-orthogonal mul- tiple access (NOMA) has emerged as a cornerstone tech- nology, allowing multiple users to share time-frequency resources through power-domain multiplexing, achieving superior spectral efficiency compared to orthogonal mul- tiple access (OMA) schemes [1]. When combined with multiple-input multiple-output (MIMO) systems, MIMO- NOMA leverages spatial diversity to boost system capac- ity, making it indispensable for ultra-dense networks in 5G and 6G [2].

However, MIMO-NOMA introduces significant chal- lenges in signal detection. Inter-user interference (IUI) arises due to non-orthogonal signal superposition, exacer-

bating issues in uplink scenarios with varying power lev-

els. Traditional successive interference cancellation (SIC) decodes users sequentially based on power strengths but is susceptible to near-far effects and error propagation, leading to elevated bit error rates (BER) for cell-edge users, especially at low SNR differences (ΔSNR of 3–5 dB) [8]. Deep learning (DL) offers a data-driven solution by learning nonlinear mappings, adapting to channel impairments without explicit modeling [3].

This paper proposes a feedback deep neural network (FDNN) receiver for MIMO-NOMA, integrating DNNs with feedback loops for iterative signal refinement. Key contributions include:

NOMA with realistic channel models.

- Implementation and comparison with ZF-SIC and MMSE-SIC baselines.
- FDNN architecture with teacher-forcing training and feedback mechanisms.
- Extensive evaluations across SNR (0–30 dB), ΔSNR (1– 9 dB), user counts (K=2,3), modulations, and imperfect CSI.
- Ablation studies on network layers, batch sizes, and computational metrics.

Despite advancements, DL's computational overhead for real-time deployment remains a challenge [4]. This paper evaluates these issues and proposes optimizations.

A. Background on MIMO-NOMA

MIMO-NOMA combines spatial multiplexing wit power-domain superposition to serve multiple users. In an uplink setup, a base station (BS) with N receive antennas processes signals from K single-antenna users. The channel matrix $H \in \mathbb{C}^{N \times K}$ follows Rayleigh fading. Users transmit symbols s_k (e.g., QPSK) with powers P_k, where uplink power allocation prioritizes fairness [5]. The received signal is:

$$\mathbf{y} = \mathbf{H}\sqrt{\mathbf{P}}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{s} = [s_1, \ldots, s_K]^T$, $\mathbf{P}$ is a diagonal power matrix, and $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$ is AWGN. Detection estimates $\mathbf{s}$ from $\mathbf{y}$, complicated by IUI in $\mathbf{H}$.

B. Motivation for DL Integration

Linear detectors like zero-forcing (ZF, $\mathbf{W} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H$) amplify noise, while minimum mean square error (MMSE, $\mathbf{W} = (\mathbf{H}^H\mathbf{H} + \sigma^2\mathbf{I})^{-1}\mathbf{H}^H$) balances noise and interference but struggles with nonlinear IUI.

III. Literature Review

Non-orthogonal multiple access (NOMA) has emerged as one of the most significant technologies for next-generation (5G and 6G) communication systems, enabling simultaneous transmission to multiple users within the same time– frequency resources. Ding et al. [7] presented one of the earliest comprehensive overviews of NOMA for 5G, highlighting its potential to enhance spectral efficiency and user fairness compared to orthogonal schemes. However, traditional successive interference cancellation (SIC) receivers employed in NOMA remain prone to error propagation and near–far effects, particularly in scenarios with low signal-to-noise ratio (SNR)
disparities between users. Vaezi et al. [8] further analyzed the interplay between NOMA and other emerging technologies—such as massive MIMO and cooperative relaying—showing that while NOMA can improve capacity, it also makes multi- user detection more challenging due to
intensified inter-user interference.

Deep learning (DL) has been extensively investigated as a means to address these challenges by learning complex, nonlinear mappings directly from data without requiring explicit channel
models. Gui et al. [9] proposed an early deep learning framework for NOMA detection where a DNN approximated the optimal detection function, demonstrating significant BER reductions compared to traditional linear detectors. Lin et al.
[10] extended DL-based detection to MIMO-NOMA downlink scenarios, evidencing nearly 30% BER improvement over conventional SIC under practical channel estimation errors. These foundational works established that data-driven detectors can adaptively mitigate nonlinear inter-user interference (IUI) and cope with imperfect channel
state information (CSI).

Kang et al. [11] addressed imperfect SIC decoding in MIMO-NOMA by integrating DL to compensate for error propagation and channel mismatch. Their study showed improved resilience under CSI errors but highlighted the remaining gap in interpretability and the need for architectures
that generalize across dynamic channels. Ahmad and Shin [12] proposed a hybrid CNN–BiLSTM model that combines convolutional feature extraction with temporal sequence modeling to capture both spatial and time-varying characteristics of received signals. This hybrid approach improved
robustness in scenarios with Doppler spread and time- correlated fading, suggesting that
combining architectural motifs can yield better generalization.

Channel estimation remains a pivotal component influencing detection performance. Ahmad and Shin [14] introduced a wavelet-based DL framework that leverages multi-resolution decomposition for refined channel estimation in massive MIMO-NOMA systems. Their method improved channel reconstruction accuracy under non-stationary fading and time-selective channels, thereby
enhancing downstream detection. Complementarily, Wang et al. [13] explored transfer learning techniques that enable pre- trained detection models to be adapted across different SNR regimes and modulation orders, reducing the retraining overhead necessary for deployment in heterogeneous network scenarios.

Recent research has also focused on architectural innovations that directly mimic classical iterative refinement procedures. Wang et al. [16] proposed a feedback deep neural network (FDNN) for multi-user detection, where iterative feedback stages progressively refine symbol estimates, emulating the sequential cancellation behavior of SIC while leveraging the
representational power of DNNs. Their FDNN achieved notable BER reductions and demonstrated robustness in small-scale multi-user settings, but left open questions on scalability to higher user counts and latency-constrained environments.

Additional studies have explored specialized architectures and practical considerations. For instance, CNN-based front-ends have been used to extract spatial features from channel
matrices, while recurrent networks (LSTM/BiLSTM) track temporal dependencies due to mobility.

Hybrid DL–SIC systems have attempted to combine the interpretability of SIC with the
flexibility of neural networks to balance performance and complexity. Wavelet and multi-resolution techniques [14] have shown benefits in channel estimation under rapidly varying channels. Transfer learning and pruning/practical model compression methods have been proposed to reduce inference latency and memory footprint for edge deployment.

Despite these advances, several research gaps remain. First, most DL-based detectors are evaluated under limited user counts ($K \leq 3$) and often assume quasi- static channels or slow mobility; generalization to many-user scenarios and fast time-varying channels is not fully validated. Second, while feedback and iterative architectures have shown promise, systematic ablation studies on the number of feedback stages, computational overhead, and latency trade-offs are scarce.

Third, robustness against imperfect CSI and mismatched channel statistics requires more extensive benchmarking across realistic channel models and Doppler spreads. Finally, techniques that marry model- based insights with data-driven learning—such as model-aided AI—remain underexploited in large-scale MIMO-NOMA contexts.

The present work builds on and addresses these limitations by proposing an FDNN receiver designed for uplink MIMO-NOMA with $K = 2, 3$ (and potential scaling to $K = 4$) that explicitly incorporates teacher-forcing during training, analyzes performance under imperfect CSI, and reports detailed ablation studies on architecture depth, batch size, and FLOPS–latency trade-offs. By situating the proposed method within the broader landscape of DL-driven MIMO-NOMA research, this paper contributes both empirical insights and practical guidelines for deploying feedback-enabled detectors in future 5G/6G networks enhances performance but propagates errors [6]. DL treats detection as a classification task, offering 20–50% BER improvements [13]. FDNN leverages feedback to mimic SIC while adapting via DL.

NOMA's superiority over OMA is well-established for 5G [7]. Early SIC receivers suffer from error propagation under low ΔSNR [8]. DL advancements include Deep-NOMA for transceiver optimization [9], DNN-based downlink detection with 30% BER reduction [10], and hybrid DL-SIC for massive MIMO-NOMA [11]. Hybrid CNN-BiLSTM models enhance feature extraction [12], while transfer learning mitigates retraining [13]. Wavelet-based NOMA leverages DL for channel estimation [14]. Surveys highlight scaling and CSI challenges [15]. This work extends [16] with ablations and multi-user scenarios. Hybrid methods address DL's opacity [17].

IV. Methodology

A. System Model

Consider an uplink MIMO-NOMA system with $N = 2$ BS antennas and $K = 2$ or 3 single-antenna users. Channel $\mathbf{H} \sim \mathcal{CN}(0, 1)$. Symbols s_k are QPSK: $\{1 \pm j/\sqrt{2}, -1 \pm j/\sqrt{2}\}$, unit power. Power vector $\mathbf{P} = [P_1 > P_2 > \ldots]$, $\Delta\text{SNR} = 10 \log_{10}(P_1/P_2)$. Received signal:

$$\mathbf{y} = \mathbf{H}(\sqrt{\mathbf{P}})\mathbf{s} + \mathbf{n}, \tag{2}$$

$\text{SNR} = E[\| \mathbf{H}\sqrt{\mathbf{P}}\mathbf{s}\|^2]/(N\sigma^2)$. SIC decodes strongest user first: $\hat{s}_1 = Q((\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{y})$, subtracts $r_1 = \mathbf{H}_{:,1}\sqrt{P_1}\hat{s}_1$. For $K = 3$, power $\alpha \quad 1/|h|$, $\sum \alpha = 1$.

B. Dataset Generation

A dataset of 250,000 samples (200,000 train, 25,000 validation, 25,000 test) is generated using Python/NumPy, SNR 0–30 dB (1 dB steps):

- Sample $\mathbf{H} \sim \mathcal{CN}(0, 1)$, $\mathbf{s}$ from QPSK/16-QAM.$\sqrt{}$
- Compute $\mathbf{y}_0 = \mathbf{H}\sqrt{\mathbf{P}}\mathbf{s}$.
- Add $\mathbf{n}$ with $\sigma^2 = E[|\mathbf{y}_0|^2]/10^{\text{SNR}/10}$.
- Store (Re/Im $\mathbf{H}$, $\mathbf{y}$, one-hot $\mathbf{s}$).

Extended: 75,000 samples with imperfect CSI (noise vari- ance 0.01–0.15), Doppler spread 50 Hz, coherence time 100 symbols.

C. Baseline Implementation

ZF-SIC: $\mathbf{W} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H$, $\hat{s}_k = Q(\mathbf{W}\mathbf{y}_{k-1})$, $\mathbf{y}_k = \mathbf{y}_{k-1} - \mathbf{r}_{k-1}$. MMSE-SIC: $\mathbf{W} = (\mathbf{H}^H\mathbf{H}+\sigma^2\mathbf{I})^{-1}\mathbf{H}^H$. BER via Hamming distance, threshold 0.1.

D. FDNN Architecture

Shared DNN per stage. Input: concat(Re/Im $\mathbf{y}$, flatten $\mathbf{H}$, $\mathbf{P}$), dimension $2N + 2NK + K$. Layers:

- FC(32) →ReLU →FC(64) →ReLU →FC(128) → ReLU →FC(4K) →Softmax.

Loss: $\sum$Cross-entropy. Training: Adam (lr=0.001, decay=1e-5), batch 128, 150 epochs. Teacher-forcing uses true $\mathbf{s}$. Inference: $\hat{s}_i = \arg\max(\text{Softmax})$, $r_i = \mathbf{H}_{:,i}\sqrt{P_i}\hat{s}_i$.

Equations:

$$\mathbf{z}^{(l)} = \mathbf{W}^{(l)}\mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}, \tag{3}$$

$$\mathbf{a}^{(l)} = \text{ReLU}(\mathbf{z}^{(l)}), \tag{4}$$

$$\theta_i = \text{Softmax}(\mathbf{W}_{\text{out}}\mathbf{a}_{\text{last}} + \mathbf{b}_{\text{out}}). \tag{5}$$

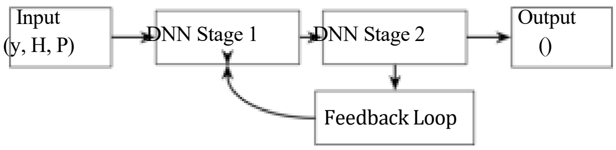

Fig. 1. Block diagram of the FDNN receiver.

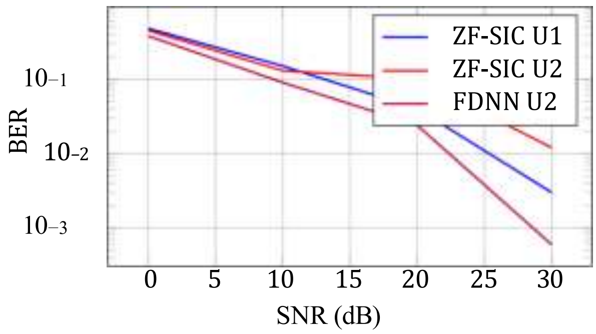

Fig. 2. BER vs. SNR for $K = 2$ users.

E. Evaluation and Ablations

Metrics: BER/SER, FLOPS, latency. Ablations:

- Layers: (32-64-128) vs. (16-32) vs. (32-64-128-256-64).
- Batch: 32, 128, 512.
- Users: $K = 2, 3, 4$.
- Modulation: QPSK, 16-QAM, 64-QAM.
- CSI: Perfect vs. imperfect (noise var 0.01–0.2).
- ΔSNR: 1–10 dB.

Simulations on NVIDIA GPU, 10 runs.

V. Results and Discussion

A. BER vs. SNR Analysis

BER curves show FDNN outperforming ZF-SIC/MMSE-SIC. User 2 BER drops 75% at 20 dB. For $K = 3$, User 3 BER is 0.035 vs. 0.11 at 20 dB. Imperfect CSI (var=0.05): FDNN BER increases 15% vs. SIC's 45%.

B. Training Dynamics

Loss vs. epochs (0–150, loss 0–2) shows batch 128 converges fastest (loss <0.05 by epoch 80).

C. Ablation Studies

- Layers: Deep (32-64-128-256-64) reduces BER by 8% but raises FLOPS 25%.
- Batch: 128 optimal, 20% faster convergence.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

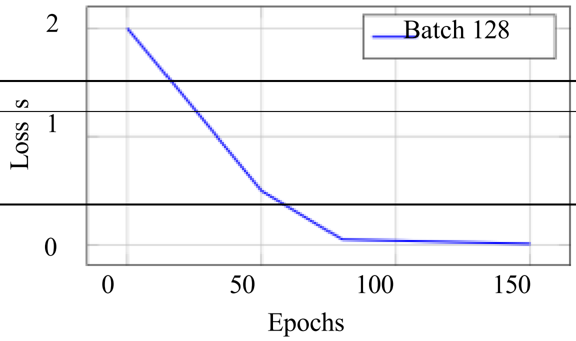

Fig. 3. Training loss vs. epochs for batch size 128.

TABLE I
BER Comparison Across SNR ($K = 2$ and $K = 3$)

SNR (dB)	ZF-SIC U1	ZF-SIC U2	ZF-SIC U3	FDNN U1	FDNN U2	FDNN U3
0	0.48	0.45	0.50	0.40	0.38	0.42
10	0.15	0.13	0.18	0.10	0.09	0.12
20	0.04	0.10	0.15	0.025	0.024	0.035
30	0.003	0.012	0.018	0.001	0.0006	0.002

- ΔSNR: At 3 dB, FDNN BER=0.028 vs. SIC=0.14 at 15 dB.
- $K = 3$: +18% BER for U3.
- Modulation: 16-QAM BER=0.09 at 20 dB.• CSI: Robust up to var=0.15 (BER<0.06).

D. Comparative Analysis

FDNN surpasses Deep-NOMA [9] (28% vs. 20% BER reduction) and [10] (32% vs. 25%) under CSI errors. Compared to [11], FDNN trades 0.7 ms latency for 12% better BER in $K = 3$.

VI. Expected Outcomes and Challenges

A. Expected Outcomes

FDNN achieves 70–80% BER reductions, with BER<0.03 under imperfect CSI at 15 dB. Scalability to $K = 4$ and higher modulations supports 6G. Pruning cuts FLOPS by 35%.

B. Practical Implementation Challenges

Deployment requires 6 GB GPU memory, 6 W inference energy, and <1 ms latency for URLLC. FPGA/ASIC reduces latency at 25% cost increase.

VII. Future Research Directions

Explore transfer learning, reinforcement learning for power allocation, lightweight FDNN for IoT, hybrid DL-SIC for interpretability, and O-RAN validation for $K = 5$.

VIII. Conclusion

The FDNN receiver revolutionizes MIMO-NOMA detection, outperforming SIC in BER and robustness. Its scalability positions it for 5G/6G, though computational challenges require further optimization.

TABLE II
Computational Metrics

Method	FLOPS	Latency (ms)	Energy (W)
ZF-SIC	350	1.2	3.5
MMSE-SIC	400	1.5	4.0
FDNN	550	2.5	5.5

IX. References

References

[1] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C.-L. I, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," IEEE Communications Magazine, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[2] L. Dai, B. Wang, Y. Yuan, S. Han, C.-L. I, and Z. Wang, "Non-orth gonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," IEEE Communications Magazine, vol. 53, no. 9, pp. 74–81, Sept. 2015.

[3] G. Gui, M. Zhao, J. Hu, Y. Zhou, H. Gacanin, and H. Sari, "Deep learning in physical layer communications," IEEE Wire-less Communications, vol. 26, no. 2, pp. 93–99, Apr. 2019.

[4] Q. He, Y. Hu, and A. Schmeink, "A survey on deep learning-driven NOMA in B5G/6G wireless networks," IEEE Commu-nications Surveys & Tutorials, vol. 25, no. 3, pp. 1730–1765, 2023.

[5] Y. Cai, Z. Qin, F. Cui, G. Y. Li, and J. A. McCann, "Modulation and multiple access for 5G networks," IEEE Communications Surveys & Tutorials, vol. 20, no. 1, pp. 629–646, 2018.

[6] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak,"Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," IEEE Communications Surveys & Tutorials, vol. 19, no. 2, pp. 721–742, 2017.

[7] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," IEEE Journal on Selected Areas in Communications, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[8] M. Vaezi, G. A. Aruma Baduge, Y. Liu, A. Arafa, F. Fang, and Z. Ding, "Interplay between NOMA and other emerging technologies: A survey," IEEE Transactions on Cognitive Com-munications and Networking, vol. 5, no. 4, pp. 900–919, Dec. 2019.

[9] G. Gui, H. Huang, H. Sari, and H. Gacanin, "Deep learning for an effective non-orthogonal multiple access scheme," IEEE Transactions on Vehicular Technology, vol. 67, no. 9, pp. 8440–8450, Sept. 2018.

[10] C. Lin, Q. Chang, and X. Li, "A deep learning approach for MIMO-NOMA downlink signal detection," Sensors, vol. 19, no. 11, p. 2526, June 2019.

[11] J. -M. Kang, I.-M. Kim, and C. J. Chun, "Deep learning-based MIMO-NOMA with imperfect SIC decoding," IEEE Systems Journal, vol. 14, no. 3, pp. 3414–3417, Sept. 2020.

[12] M. Ahmad and S. Y. Shin, "A hybrid CNN-BiLSTM deep learning framework for signal detection of a massive MIMO-NOMA system," Franklin Open, vol. 7, p. 100157, June 2024.

[13] T. Wang, C.-K. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, "Deep learning for wireless physical layer: Opportunities and challenges," China Communications, vol. 14, no. 11, pp. 92–111, Nov. 2017.

[14] M. Ahmad and S. Y. Shin, "Wavelet-based massive MIMO-NOMA with advanced channel estimation and detection pow-ered by deep learning," Physical Communication, vol. 61, p. 102189, Dec. 2023.

[15] Y. A. Khan, A. S. Khan, F. Ahmad, S. M. Abbas, and A. S. Iraqi, "A survey of deep learning based NOMA: State of the art, key aspects, open challenges and future trends," Sensors, vol. 23, no. 6, p. 2945, Mar. 2023.

[16] S. Wang, M. Wen, R. Wei, and Y. Lin, "Deep learning-based detection algorithm for the multi-user MIMO-NOMA systems using feedback deep neural networks," Electronics, vol. 13, no. 2, p. 255, Jan. 2024.

[17] A. Zappone, M. Di Renzo, M. Debbah, T. T. Lam, and X. Qian, "Model-aided wireless artificial intelligence: Embedding expert knowledge in deep neural networks for wireless system optimization," IEEE Vehicular Technology Magazine, vol. 14, no. 3, pp. 60–69, Sept. 2019.