Biometric Identification Through Auditory EEG Signatures

Ly Ngoc Vu
Faculty of Information Technology
Industrial University of Ho Chi Minh City, Vietnam
dezzhuge@gmail.com

Bui Thi Ngoc Tran
Faculty of Information Technology
Industrial University of Ho Chi Minh City, Vietnam
tranbui2907@gmail.com

Huynh Cong Bang
Faculty of Information Technology
Industrial University of Ho Chi Minh City, Vietnam
bang_01036047@iuh.edu.vn

Nguyen Minh Hanh
Faculty of Information Technology
Industrial University of Ho Chi Minh City, Vietnam
hanh_01036046@iuh.edu.vn

***Abstract*—Intelligent authentication systems for real-world deployment must be robust to noise, scalable to new users, and capable of operating under open-set conditions. However, many existing approaches remain constrained by closed-set assumptions and limited resilience to signal degradation, reducing their practical applicability. This research presents a noise-resilient open-set authentication framework that systematically addresses both adaptive signal enhancement and discriminative representation learning. A WaveNet-based denoising module, optimized using a scale-invariant signal-to-noise ratio objective, mitigates environmental and physiological noise. Subsequently, a deep metric learning module built on ResNet backbones and angular margin-based losses learns a compact embedding space for reliable identity discrimination and unseen-user detection. Using brain-derived signals as a case study, the proposed framework achieves high identity retrieval performance (Precision@1 up to 98%) and robust open-set verification (TAR@1%FAR $\approx$ 95%). Further analysis highlights the importance of explicitly decoupling noise mitigation and embedding learning objectives in intelligent authentication systems. Overall, the proposed framework provides a robust solution for open-set scenarios.**

***Index Terms*—Biometric identification, EEG authentication, Open-set recognition, Deep metric learning, Signal enhancement.**

I. Introduction

The rapid proliferation of intelligent systems in security-sensitive applications has intensified the demand for authentication mechanisms that are not only accurate but also robust, scalable, and adaptable to real-world conditions. Modern authentication systems are increasingly deployed in dynamic environments such as smart devices, wearable technologies, and cyber-physical systems, where signal quality, user populations, and operating conditions cannot be strictly controlled. As a result, intelligent authentication solutions must handle noisy inputs, accommodate newly encountered users, and operate reliably in open-set scenarios [1]–[3].

Although deep learning has improved biometric performance, traditional systems often fail under open-set conditions and signal degradation. Most existing approaches rely on closed-set assumptions unrealistic for scalable deployment [1]–[3], while real-world data remains vulnerable to physiological and environmental noise [1], [4].

To address these challenges, representation learning and deep metric learning have emerged as powerful paradigms for intelligent authentication. By learning discriminative embedding spaces that preserve semantic similarity, metric learning enables identity verification, retrieval, and open-set recognition within a unified framework [5]–[8]. However, the effectiveness of embedding-based authentication systems is highly sensitive to the quality of the input signal, underscoring the need for robust signal enhancement mechanisms [1], [4].

In parallel, recent advances in deep neural architectures for signal enhancement have demonstrated strong potential to mitigate noise in complex, non-stationary signals. Models such as WaveNet have shown remarkable ability to learn signal structures directly from raw data, enabling adaptive denoising across diverse noise conditions [14]. Despite these advances, integrating noise-adaptive signal enhancement with discriminative representation learning for open-set authentication remains underexplored, particularly from a system-level perspective [1], [2].

This research addresses these gaps by proposing an intelligent authentication framework that systematically addresses noise resilience and open-set recognition. The framework integrates a WaveNet-based adaptive denoising module with a deep metric-learning architecture built on ResNet backbones and angular-margin–based loss functions [7], [8], [11]. By combining noise-aware signal enhancement with embedding-level discrimination, the proposed system is designed to operate reliably in realistic environments and scale to previously unseen users [1], [5], [6]. Using brain-derived signals as a representative case study, extensive experiments demonstrate that the proposed approach achieves high robustness, high-precision identity retrieval, and strong open-set verification performance (TAR at low FAR) [1], [2], [4]. The results highlight the importance of explicitly decoupling noise mitigation and representation learning objectives when designing deployable intelligent authentication systems [7], [8].

II. Related work

Intelligent authentication systems have attracted growing attention as security requirements increase in dynamic and resource-constrained environments such as smart devices, wearable platforms, and cyber–physical systems. Traditional authentication mechanisms, including passwords and token-based methods, are prone to well-known security vulnerabilities. At the same time, conventional biometric systems based on facial images, fingerprints, or voice signals remain susceptible to spoofing attacks [1]. As a result, learning-based authentication approaches have been widely investigated to enhance robustness and adaptability.

Early EEG authentication studies primarily focused on closed-set classification using recurrent or convolutional architectures [3]. While effective in controlled settings, these models struggle with unseen users. To address this, metric learning approaches (e.g., contrastive and triplet losses) have been adopted to learn distance-based embeddings [5], [6]. Recently, angular margin–based losses like ArcFace have demonstrated superior separability [7], yet their resilience to non-stationary noise remains limited.

Robustness to noise is a critical challenge in practical authentication systems, as real-world signals are often corrupted by environmental interference and physiological artifacts [4]. Traditional filtering and handcrafted feature extraction methods have been widely used to mitigate noise, but their performance is limited under complex, non-stationary conditions. Recent deep learning–based denoising approaches, particularly those capable of modeling temporal dependencies, have shown strong potential for adaptive noise suppression. Despite these advances, most existing studies treat signal enhancement and representation learning as independent components, without explicitly optimizing their interaction for open-set authentication.

Brain-derived signals have also been explored as an alternative authentication modality due to their inherent uniqueness and resistance to spoofing. Prior studies have reported promising identification performance using both traditional machine learning and deep learning models [1]–[3]. However, these systems remain vulnerable to noise, session variability, and deployment constraints, and open-set operation is often insufficiently addressed [4]. Additionally, ethical considerations around privacy and user consent are crucial when deploying brain-derived signal systems. Acknowledging these concerns can build trust with users by ensuring their data is handled with care, transparency, and in compliance with relevant regulations. By addressing these ethical aspects, the potential of brain-derived signals as a secure and reliable modality can be fully realized, anticipating stakeholders' concerns while reinforcing their value in authentication.

In summary, although prior work has demonstrated the effectiveness of metric learning and noise suppression for authentication, there remains a lack of integrated frameworks that jointly address noise resilience, discriminative representation learning, and open-set recognition at the system level. This gap motivates the proposed approach. Unlike prior EEG biometric studies that rely on single-stage classification models, this work formulates EEG authentication as a two-stage learning problem, decoupling deep representation learning from identity discrimination via metric learning [5]–[8]. This design enables robust open-set authentication and superior generalization to unseen subjects.

III. Methodology

This study proposes a two-stage authentication framework that enhances robustness and open-set capability by explicitly decoupling noise mitigation from representation learning. The overall architecture consists of a noise-adaptive signal enhancement stage followed by a deep metric learning–based embedding stage, as illustrated in Fig. 1.

A. Stage I: Noise-Adaptive Signal Enhancement

The first stage aims to mitigate environmental and physiological noise that commonly degrades real-world signals [4]. A WaveNet-based denoising model is employed to learn a direct mapping from noisy inputs to cleaner signal representations [14]. The model is trained using a scale-invariant signal-to-noise ratio (SI-SNR) objective, which encourages faithful signal reconstruction while remaining insensitive to amplitude variations. By explicitly performing denoising before representation learning, this stage reduces noise-induced distortions and improves signal consistency across sessions and operating conditions. Fig. 2

Noise synthesis and denoiser training. To train Stage I, we construct paired samples $(x_{\text{noisy}}, x_{\text{clean}})$ using only subjects in $\mathcal{S}_{\text{train}}$. We define x_{clean} as the pseudo-clean EEG from the dataset and generate $x_{\text{noisy}} = x_{\text{clean}} + \alpha n$, where n is sampled from Gaussian / power-line / EMG noise and α is set to match SNR levels in $\{0, 5, 10, 20\}$ dB. Care is taken to ensure no information leakage between training and evaluation splits during noise mixing. The WaveNet denoiser is optimized with the SI-SNR objective to recover x_{clean} from x_{noisy}.

B. Stage II: Discriminative Representation Learning

In the second stage, denoised signals are processed by a deep metric learning model to learn identity-discriminative embeddings [5], [6]. Residual convolutional neural networks (ResNet) are used as backbone architectures due to their stability and strong representation capacity [11]. Angular margin–based loss functions are employed to enforce compact intra-class clustering and clear inter-class separation in the embedding space [7]. In addition, general pair-weighted metric objectives (e.g., multi-similarity loss) can further improve embedding discrimination by leveraging multiple positive and negative relationships within a batch [8]. This embedding-based formulation naturally supports open-set recognition, enabling the detection of unseen users based on distance to learned identity prototypes in the embedding space [5], [6].

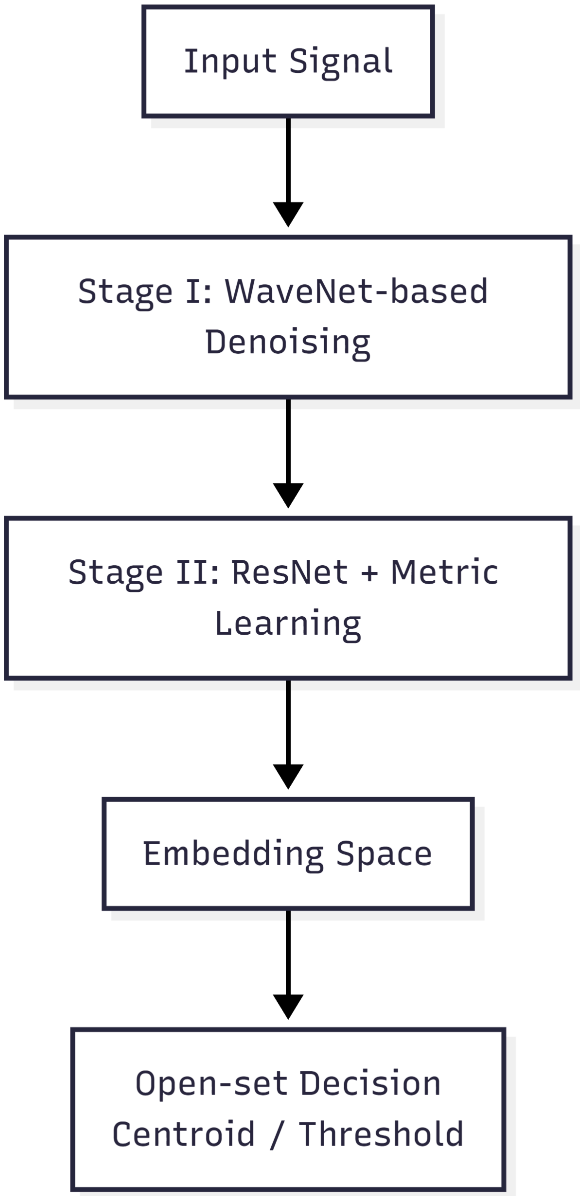

Fig. 1. Overview of the proposed two-stage authentication framework.

IV. Experimental Setup

Dataset and Segmentation. We evaluate our framework on the *Auditory Evoked Potential EEG–Biometric Dataset* [4], which contains EEG recordings from 20 participants. Each subject participated in 12 sessions, with each session lasting approximately 2 minutes. To generate training and evaluation samples, continuous EEG signals were segmented into 4-second windows with a 2-second overlap (50% hop size). We refer to each segmented window as a *query epoch* (x). This segmentation strategy increases the number of samples while preserving temporal continuity.

Subject Partitioning. To evaluate open-set generalization, we partition the 20 subjects into two disjoint sets:

- **Known Subjects ($\mathcal{S}_{\textbf{known}}$)**: 16 subjects (80% of the population) are used to train the model and form the enrollment database.
- **Unknown Subjects ($\mathcal{S}_{\textbf{unknown}}$)**: 4 subjects (20% of the population) are strictly held out to represent unseen impostors during testing. Since $\mathcal{S}_{\text{known}}$ and $\mathcal{S}_{\text{unknown}}$ are disjoint by design, any query from these 4 subjects should be rejected by the system.

Data Partitioning for Known Subjects. For the 16 subjects in $\mathcal{S}_{\text{known}}$, we split their epochs into three subsets using a stratified shuffle split:

- **Training Split ($\mathcal{S}_{\textbf{train}}$)**: 64% of the epochs. Used to train the model and compute enrollment prototypes.
- **Validation Split ($\mathcal{S}_{\textbf{val}}$)**: 16% of the epochs. Used for hyperparameter tuning and to calibrate the rejection threshold τ.
- **Test Split ($\mathcal{S}_{\textbf{test}}$)**: 20% of the epochs. Used as genuine queries (positive samples) for evaluation.

Enrollment and Decision Rule. We define the enrollment set $\mathcal{S}_{\text{enroll}}$ as the set of identity prototypes derived from $\mathcal{S}_{\text{known}}$. For each known subject $c \in \mathcal{S}_{\text{known}}$, we compute a prototype μ_c by averaging the embeddings of their training epochs: $\mu_c = \frac{1}{|\mathcal{E}_c|}\sum_{x\in\mathcal{E}_c} f(x)$, where $\mathcal{E}_c \subset \mathcal{S}_{\text{train}}$. For an incoming query epoch x, the system computes its embedding $z = f(x)$ and the distance to the nearest enrolled prototype: $d^* = \min_{c\in\mathcal{S}_{\text{enroll}}} \|z - \mu_c\|_2$. The identity is accepted if $d^* \leq \tau$, otherwise it is rejected as unknown.

Metrics. We evaluate performance using precision for closed-set identification and True Acceptance Rate (TAR) for open-set verification.

- **Precision@1 (P@1)**: Measures the accuracy of identifying genuine users from $\mathcal{S}_{\text{test}}$ against the enrolled set $\mathcal{S}_{\text{enroll}}$.

$$\text{P@1} = \frac{1}{N}\sum_{x\in\mathcal{S}_{\text{test}}} \mathbb{I}(\text{id}(x) = \arg\min_{c\in\mathcal{S}_{\text{enroll}}} \|f(x) - \mu_c\|_2)$$

- **True Acceptance Rate (TAR)**: Measures the fraction of genuine queries from $\mathcal{S}_{\text{test}}$ correctly accepted at a fixed threshold τ.

$$\text{TAR}(\tau) = \frac{|\{x \in \mathcal{S}_{\text{test}} : d^*(x) \leq \tau\}|}{|\mathcal{S}_{\text{test}}|}$$

The threshold τ is set to satisfy a False Acceptance Rate (FAR) of 1%, where FAR is calculated using impostor queries from $\mathcal{S}_{\text{unknown}}$:

$$\text{FAR}(\tau) = \frac{|\{x \in \mathcal{S}_{\text{unknown}} : d^*(x) \leq \tau\}|}{|\mathcal{S}_{\text{unknown}}|}$$

Baselines. We compare our Two-Stage framework (WaveNet Denoiser + ResNet Metric Learning) against: 1) **Single-Stage ResNet**: The same ResNet backbone trained directly on noisy data without enhancement. 2) **LSTM/BiLSTM**: Recurrent neural networks often used for EEG time-series classification [9], [10]. All models are trained and evaluated on the same splits.

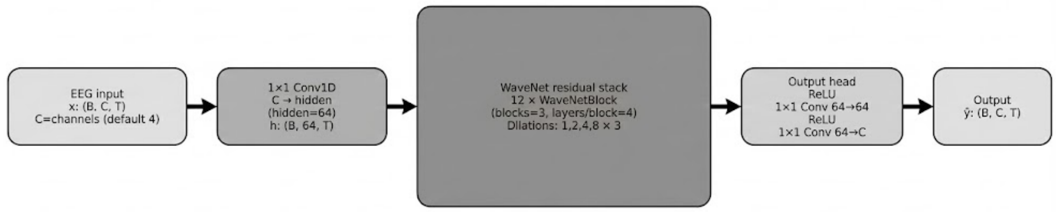

Fig. 2. WaveNetDenoiser architecture for EEG denoising.

V. Results and Discussion

Experimental results demonstrate that the proposed two-stage framework significantly outperforms both recurrent and single-stage convolutional baselines, particularly under noisy conditions. Table I summarizes the performance comparison across all baseline methods under three noise types.

Comparison with Baseline Methods. We compare our proposed two-stage framework against three baseline approaches: LSTM, BiLSTM, and Single-Stage ResNet-34 (trained directly on noisy data without enhancement). As shown in Table I, the recurrent baselines (LSTM and BiLSTM) achieve reasonable accuracy in low-noise scenarios but their performance drops sharply as noise complexity increases. Under EMG noise, LSTM achieves P@1 of 87.10% and BiLSTM achieves 88.50%, while our proposed method achieves 97.62%—an improvement of over 9 percentage points. Fig. 3 further illustrates the performance gap between recurrent baselines.

The Single-Stage ResNet-34 (without denoising) provides a stronger baseline, achieving P@1 of 96.66% under EMG noise. By integrating the WaveNet denoiser (Two-Stage), performance improves to 97.62% (+0.96%). More importantly, open-set robustness (TAR@1%FAR) improves from 95.25% to 95.51%. This improvement is consistent across all noise types, validating that explicit signal enhancement helps the metric learning model generate more discriminative and stable embeddings.

Fig. 3. Performance comparison of single-stage recurrent baselines (LSTM and BiLSTM) showing accuracy drop under increasing noise complexity.

TABLE I
Comparison with Baseline Methods (P@1 and TAR@1%FAR)

Noise Condition	Method	P@1 (%)	TAR@1%FAR (%)
Gaussian	LSTM (Single-stage)	89.21	88.45
	BiLSTM (Single-stage)	90.15	89.12
	ResNet-34 (Single-stage)	93.31	95.38
	Proposed (Two-stage)	**93.80**	**95.37**
Power-line	LSTM (Single-stage)	85.40	82.10
	BiLSTM (Single-stage)	86.75	83.50
	ResNet-34 (Single-stage)	96.92	95.38
	Proposed (Two-stage)	**98.15**	**95.64**
EMG	LSTM (Single-stage)	87.10	85.30
	BiLSTM (Single-stage)	88.50	86.20
	ResNet-34 (Single-stage)	96.66	95.25
	Proposed (Two-stage)	**97.62**	**95.51**

A. Effect of Metric Learning Objectives

Table II compares different deep metric learning objectives on ResNet backbones. Across architectures, metric learning losses significantly improve discriminative embeddings compared with conventional closed-set formulations by explicitly optimizing distances in the embedding space [5], [6]. In particular, multi-similarity loss and ArcFace achieve the strongest performance, indicating better intra-class compactness and inter-class separability [7], [8]. ResNet backbones are used due to their stable optimization and strong representational capacity [11].

TABLE II
Performance comparison of deep metric learning losses on ResNet backbones.

Backbone	Loss Function	Retrieval (P@1, %)
ResNet-18	Triplet Loss	97.89
ResNet-18	Contrastive Loss	97.96
ResNet-18	Multi-Similarity	98.67
ResNet-18	ArcFace	98.60
ResNet-34	Triplet Loss	94.27
ResNet-34	Contrastive Loss	98.07
ResNet-34	Multi-Similarity	**98.84**
ResNet-34	ArcFace	98.81
ResNet-50	Triplet Loss	92.76
ResNet-50	Contrastive Loss	97.65
ResNet-50	Multi-Similarity	98.52
ResNet-50	ArcFace	98.56

Overall, the proposed approach maintains high identity retrieval performance (P@1) and achieves strong open-set verification performance, demonstrating robust generalization beyond the closed-set assumption commonly adopted in earlier EEG biometric systems [1]–[3].

B. Robustness Under Noise and Ablation Study

Table III reports P@1 and SI-SNR across different noise types. The results indicate that the framework remains stable under Gaussian, power-line, and EMG noise, suggesting that explicit enhancement improves downstream embedding discrimination. (Note that SI-SNR is included as a reconstruction-oriented indicator of denoising quality [13].)

TABLE III
PERFORMANCE OF THE PROPOSED TWO-STAGE FRAMEWORK UNDER DIFFERENT NOISE CONDITIONS.

Noise	Model	P@1 (%)	SI-SNR (dB)
Gaussian	ResNet-34 + Multi-Similarity	93.80	12.62
Gaussian	ResNet-18 + Multi-Similarity	**94.24**	12.61
Gaussian	ResNet-34 + ArcFace	93.36	12.69
Power-line	ResNet-34 + Multi-Similarity	98.15	35.55
Power-line	ResNet-18 + Multi-Similarity	**98.42**	34.09
Power-line	ResNet-34 + ArcFace	98.06	35.89
EMG	ResNet-34 + Multi-Similarity	**97.62**	14.18
EMG	ResNet-18 + Multi-Similarity	96.26	14.17
EMG	ResNet-34 + ArcFace	96.44	14.09

To quantify the contribution of Stage I, Table IV presents an ablation study comparing single-stage training versus the full two-stage pipeline under each noise condition. Removing the noise-adaptive enhancement module results in a consistent drop in P@1, indicating that denoising improves embedding stability and helps preserve the discriminative structure for unseen-user detection. These findings support the hypothesis that explicitly separating noise mitigation from representation learning yields a more robust and deployable authentication pipeline.

TABLE IV
ABLATION STUDY OF THE PROPOSED TWO-STAGE FRAMEWORK UNDER DIFFERENT NOISE CONDITIONS.

Noise	Framework	Model	P@1 (%)	TAR@1%FAR (%)
Gaussian	Single-stage	ResNet-34 + Multi-Similarity	93.31	95.38
Gaussian	Two-stage	ResNet-34 + Multi-Similarity	**93.80**	94.37
Power-line	Single-stage	ResNet-34 + Multi-Similarity	96.92	95.38
Power-line	Two-stage	ResNet-34 + Multi-Similarity	**98.15**	95.64
EMG	Single-stage	ResNet-34 + Multi-Similarity	96.66	95.25
EMG	Two-stage	ResNet-34 + Multi-Similarity	**97.62**	95.51

From a system-level perspective, the proposed two-stage design enables targeted optimization of enhancement and embedding objectives, improving interpretability and robustness in realistic conditions. This is particularly important for EEG-based authentication, where session variability and non-stationary noise frequently degrade performance [1], [2], [4].

VI. CONCLUSION

This research presented a two-stage intelligent authentication framework that integrates adaptive noise mitigation with deep metric learning-based representation learning. By explicitly decoupling denoising from identity discrimination, the proposed approach enhances robustness, scalability, and open-set recognition. Experimental results demonstrate that the framework outperforms conventional single-stage deep learning models, achieving strong performance in identity retrieval and unseen-user detection. The proposed design offers a practical and deployable solution for secure authentication and can be extended to other intelligent sensing and biometric applications.

VII. FUTURE WORK

Two directions are promising for future research. First, we will expand the evaluation to larger, more diverse EEG datasets (e.g., more subjects, more extended time spans, additional acquisition devices, and environments) to better assess cross-session and cross-device generalization. Second, we will investigate more advanced denoising architectures beyond WaveNet, such as state-space sequence models (e.g., Mamba layers), to improve robustness under non-stationary artifacts and real-world noise, and to study whether end-to-end optimization further benefits open-set authentication.

REFERENCES

[1] M. Fallahi, T. Strufe, and P. Arias-Cabarcos, “BrainNet: Improving Brainwave-Based Biometric Recognition with Siamese Networks,” in *Proc. IEEE PerCom*, 2023, pp. 53–60.

[2] H. Yang, W. Zheng, Y. Zhang, and B. L. Lu, “EEG Authentication Based on Deep Learning of Triplet Loss,” *Neural Network World*, vol. 32, no. 5, pp. 269–283, 2022.

[3] X. Zhang, L. Yao, S. S. Kanhere, Y. Liu, T. Gu, and K. Chen, “MindID: Person Identification from Brain Waves through Attention-Based Recurrent Neural Network,” arXiv:1711.06149, 2017.

[4] N. A. Alzahab *et al.*, “Auditory Evoked Potential Electroencephalography-Biometric Dataset,” *Data in Brief*, vol. 57, p. 111065, 2024.

[5] R. Hadsell, S. Chopra, and Y. LeCun, “Dimensionality Reduction by Learning an Invariant Mapping,” in *Proc. IEEE CVPR*, 2006, pp. 1735–1742.

[6] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering,” in *Proc. IEEE CVPR*, 2015, pp. 815–823.

[7] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “ArcFace: Additive Angular Margin Loss for Deep Face Recognition,” in *Proc. IEEE/CVF CVPR*, 2019, pp. 4690–4699.

[8] X. Wang *et al.*, “Multi-Similarity Loss with General Pair Weighting for Deep Metric Learning,” in *Proc. IEEE/CVF CVPR*, 2019, pp. 5022–5030.

[9] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[10] A. Graves and J. Schmidhuber, “Framewise Phoneme Classification with Bidirectional LSTM,” *Neural Networks*, vol. 18, no. 5–6, pp. 602–610, 2005.

[11] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *Proc. IEEE CVPR*, 2016, pp. 770–778.

[12] D. L. Davies and D. W. Bouldin, “A Cluster Separation Measure,” *IEEE Trans. PAMI*, vol. 1, no. 2, pp. 224–227, 1979.

[13] Y. Luo and N. Mesgarani, “TasNet: Time-Domain Audio Separation Network for Real-Time, Single-Channel Speech Separation,” in *Proc. IEEE ICASSP*, 2018, pp. 696–700, doi: 10.1109/ICASSP.2018.8462116.

[14] A. van den Oord *et al.*, “WaveNet: A Generative Model for Raw Audio,” arXiv:1609.03499, 2016.