I.J. Image, Graphics and Signal Processing, **2026, 1, 70-87**
Published Online on February 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2026.01.05

Segment Wise EEG Signal Compression Using LSTM Auto Encoder for Enhanced Efficiency

Uma. M.
Department of Computational Intelligence, SRM Institute of Science and Technology, Kattankulathur, Chennai 603203, India
Email: umam@srmist.edu.in
ORCID iD: https://orcid.org/0000-0003-0976-7411

Mohammed Javidh S.
Department of Computational Intelligence, SRM Institute of Science and Technology, Kattankulathur, Chennai 603203, India
Email: mm1632@srmist.edu.in
ORCID iD: https://orcid.org/0009-0001-2631-664X

Ruchi Shah
Department of Computational Intelligence, SRM Institute of Science and Technology, Kattankulathur, Chennai 603203, India
Email: rb4726@srmist.edu.in
ORCID iD: https://orcid.org/0009-0005-4115-107X

Prabhu Sethuramalingam*
Department of Mechanical Engineering, SRM Institute of Science and Technology, Kattankullathur, Chennai-603203, India
Email: prabhus@srmist.edu.in
ORCID iD: https://orcid.org/0000-0003-0707-2720
***Correspondence Author

M. M. Reddy
Department of Mechanical Engineering, Curtin University, Miri, CDT 250, Malaysia
Email: mohan.m@curtin.edu.my
ORCID iD: https://orcid.org/0000-0002-0355-854X

Received: 22 June, 2025; Revised: 23 September, 2025; Accepted: 18 December, 2025; Published: 08 February, 2026

Abstract: Efficient compression of electroencephalogram (EEG) signals is crucial for enabling real-time monitoring, storage, and transmission in various medical and non-medical applications. This paper presents a segment-wise processing approach using temporal modeling-based auto encoders for EEG signal compression. By leveraging models such as Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Recurrent Neural Network (RNN), and Self-Attention, the proposed method effectively captures temporal dependencies in the EEG data. Segment-wise processing not only enhances compression efficiency but also significantly reduces the processing time of these sequence models. Extensive experiments demonstrate that GRU-based auto encoders offer the best performance, particularly at lower Data Reduction Factors (DRFs), achieving a minimal signal loss of 0.2% at a 50% compression ratio, making it suitable for medical applications. For non-medical scenarios, a higher compression ratio of 75% with a signal loss of 5.4% is found to be acceptable. The results indicate that the proposed approach achieves a favorable balance between compression efficiency, signal fidelity, and computational performance.

Index Terms: EEG Signal Compression; LSTM Auto encoder; Segment-Wise Processing; Biomedical Signal Processing

1. Introduction

Electroencephalography (EEG) is a widely used modality for monitoring brain activity, valued for its non-invasiveness and millisecond-scale temporal resolution. Modern EEG recordings often involve high sampling rates (250–1000 Hz or more) and dozens to hundreds of channels. As a result, raw EEG data accumulates rapidly – continuous monitoring can produce gigabytes of data in hours, and large EEG databases reach terabyte scales. This data deluge poses challenges for the storage, transmission, and real-time processing of EEG signals. Bandwidth limitations in wireless EEG devices and memory constraints in wearable or implantable systems further exacerbate the need for efficient EEG data compression. Effective compression techniques are essential to reduce data size while preserving the information content critical for clinical diagnosis (e.g., epilepsy detection) and BCI applications.

Traditional lossless compression methods (e.g. Huffman coding, run-length encoding) guarantee exact signal reconstruction but typically achieve only modest size reduction on EEG signals. For instance, general-purpose lossless algorithms often yield compression ratios around 1.5:1 to 3:1 (50–67% size reduction) on multichannel EEG. While lossless approaches ensure no diagnostic information is lost their limited compression capability motivates exploring lossy compression for greater efficiency. Lossy techniques like transform coding (using Discrete Cosine Transform (DCT) or Discrete Wavelet Transform (DWT)) can significantly reduce data volume by discarding perceptually or statistically "redundant" components of the signal. These methods exploit the fact that EEG signals have inherent spatiotemporal redundancies – adjacent time samples and neighboring channels are correlated – and thus can be represented in a more compact form. However, traditional lossy compression may introduce distortion that could obscure clinically important details if not carefully controlled. Furthermore, many existing feature-extraction approaches for EEG (such as selecting spectral or spatial features) are task-specific, meaning they are tuned to particular applications (seizure detection, BCI control, etc.) and may not generalize well to other tasks. This lack of generalization in turn limits the reusability of compressed EEG data – compression optimized for one purpose might fail to preserve information needed for a different analysis. There is a clear need for an EEG compression strategy that yields good compression ratios while preserving information in a task-independent manner. If the compressed representation itself could serve as a general feature embedding of the EEG, one could perform various downstream tasks (like classification of cognitive states or signal reconstruction for clinician review) directly on the compressed data without reverting to the full signal. Such an approach would transform compression from a mere data reduction tool into an integral part of the EEG analysis pipeline, enabling efficient data handling and multi-purpose use of stored signals. Recent advances in machine learning, particularly deep neural networks, provide a promising avenue to learn such representations. Autoencoders and sequence models can be trained to capture complex patterns in EEG, potentially outperforming hand-crafted compression techniques by learning an encoding optimized for preserving brain-signal features.

1.1. Related research work

Early research on EEG compression predominantly focused on lossless compression techniques grounded in information theory and signal coding principles. Over time, approaches evolved to include lossy transform-based methods, hybrid schemes combining lossy and lossless stages, and more recently, compressive sensing and deep learning-based methods. In this section, we provide an overview of these categories, comparing their advantages and limitations in the context of EEG data.

Lossless EEG Compression Techniques: Lossless methods ensure that the EEG signal can be perfectly reconstructed from the compressed data, which is crucial for preserving every detail of brain signals in sensitive clinical applications. Traditional lossless compressors applied to EEG include general algorithms like Huffman coding and Arithmetic Coding, which encode data statistically based on symbol frequencies. For instance, G.Antoniol et al. [1] applied Huffman coding on EEG after simple preprocessing (such as taking first-order differences of samples) and achieved compression ratios around 1.7:1 to 2.4:1 (approximately 49–58% size reduction). Their approach demonstrated that exploiting the redundancy in successive EEG samples (via derivative computation) significantly improved compression over straight entropy coding. While conventional lossless algorithms are fast and guarantee no information loss, a known limitation is their relatively low compression ratios on physiological signals. EEG signals contain significant randomness and noise alongside their informative patterns; thus, purely lossless compression can rarely shrink data beyond ~50% without discarding some information. More advanced lossless techniques have been proposed to push these limits. The advantage of lossless compression is complete fidelity – the reconstructed EEG is bit-wise identical to the original, ensuring no degradation in signal quality. This is particularly important in medical contexts or when subtle EEG details (such as slight waveform morphologies or precise amplitudes) are needed for diagnosis. However, the downside is that even the best lossless algorithms yield only moderate size reduction and may not sufficiently alleviate storage/transmission burdens for large EEG datasets.

Lossy Compression Techniques: Lossy compression exploits the tolerance for some error in the reconstructed signal in exchange for much higher compression. For EEG, a common lossy approach is transform coding: the EEG signal is transformed into a domain where its energy is more concentrated (e.g., frequency domain), and then small or high-frequency components are discarded or coarsely quantized. Discrete Cosine Transform (DCT) and Discrete

Wavelet Transform (DWT) are two widely used transforms in this regard. DCT, known as image compression (JPEG), has also been applied to EEG because it compacts signal energy into the first few coefficients, and fast algorithms exist for its computation. However, EEG signals have a broad frequency spectrum without a very sharp cutoff; significant information can be present even in what might appear as minor components. Antoniol et al. [1] observed that truncating the EEG's DCT coefficients led to relatively poor compression performance compared to other methods since the discarded components still carried non-negligible power. In other words, naive DCT-based EEG compression can cause important subtle signal features to be lost, resulting in higher distortion (or higher Percent Root-mean-square Difference, PRD) for a given compression ratio.

Wavelet transforms (DWT), on the other hand, represent EEG in a multi-resolution time-frequency space and have been found more effective in many cases. By choosing an appropriate wavelet basis, EEG's transient features (spikes, sharp waves) and oscillatory rhythms can be sparsely represented. Many lossy EEG compressors use wavelet coefficient thresholding: coefficients below a threshold are set to zero (introducing loss) and then the sparse coefficient set is encoded. Another approach is to treat the EEG channels as images or 2D matrices. Srinivasan et al. [2] proposed arranging EEG data in a 2D matrix (for example, time samples along one dimension and channel index along the other) and then applying 2D image compression techniques. In their method, a wavelet-based image coder (SPIHT – Set Partitioning in Hierarchical Trees) was applied to the EEG matrix as a lossy stage, leveraging both temporal and spatial (inter-channel) correlations. This yielded better compression than 1D methods on the same data because the 2D representation made redundancy more explicit. As shown in their results, the 2D SPIHT approach on a multichannel EEG yielded compression ratios around 1.8–2.2:1 on various datasets, whereas a comparable 1D SPIHT on each channel achieved only about 1.3–1.7:1. The improvement underscores how spatial redundancy (similar patterns across channels) can be utilized in compression – an aspect plain 1D methods ignore.

Beyond transforms, other lossy strategies include predictive coding (using linear predictors or neural network predictors to estimate samples and encode the small errors) and dictionary methods. For example, linear predictive coding was adapted for EEG by modeling each sample as a linear combination of past samples and encoding the prediction error. This can be effective if the EEG has a quasi-periodic structure. However, EEG's non-stationarity often limits simple predictors. Nonlinear predictors like neural networks have been investigated – e.g., recurrent neural network (RNN) predictors were tested in the 1990s for EEG compression. These predictors can learn to capture more complex temporal patterns, reducing error variance, but at the cost of higher computational complexity. Generally, lossy techniques achieve much higher compression than lossless ones; compression ratios of 5:1 or more are attainable depending on acceptable distortion. The limitation of lossy compression is potential information loss: if too much is discarded, the diagnostic or interpretive value of EEG can be compromised. Therefore, a critical aspect is balancing compression ratio and signal quality. Many works report metrics like PRD or root-mean-square error alongside compression ratio to quantify this trade-off.

Hybrid Compression Techniques (Lossy + Lossless): To exploit the strengths of both approaches, many researchers have designed hybrid compression systems that apply a lossy compression stage followed by a lossless stage. The lossy stage greatly reduces redundancy (e.g., zeros out small coefficients), and the lossless stage then efficiently encodes the remaining data without further information loss. The rationale is that the lossy step removes the least important information to attain a target quality, and then lossless coding squeezes the data as much as possible without sacrificing what remains. Yousri et al. [3] provide a representative example: they experimented with combinations like DWT+Arithmetic Coding, DWT+RLE, DCT+Arithmetic, and DCT+RLE (Run-Length Encoding) for EEG compression. Their findings showed that using DCT for the lossy transform and RLE for the lossless encoding yielded the best overall performance in terms of compression ratio and computational complexity. In particular, the DCT+RLE scheme achieved compression rates as high as 94% (i.e., compressed size only 6% of the original, which is roughly a 16.7:1 compression ratio) with a low reconstruction error (RMSE ≈ 0.188). Even at an extreme compression of 94%, the reconstructed EEG retained the essential shape, suffering only minor distortion. At a more moderate compression level (60% reduction, or compressed to 40% of the original size), the error was very low (RMSE ~0.065), resulting in an almost indistinguishable reconstruction. This demonstrates that a well-chosen hybrid method can dramatically shrink data volume while keeping errors within acceptable bounds. Gürkan et al [7]. Introduced an EEG compression method using CSEVS, which applies k-means clustering for efficient signal representation. The model achieves high compression ratios while preserving diagnostic information, making it a promising hybrid technique.

Deep Learning-Based EEG Compression: An autoencoder is a neural network that learns to encode the input into a smaller-dimensional latent vector and then decode it back to the original; when trained on EEG data, the encoder can be viewed as the compression function and the decoder as the decompression. Unlike fixed transforms, an autoencoder can potentially learn complex nonlinear features that are optimally suited for representing EEG signals with minimal error. Various forms of autoencoders have been tested: e.g., sparse autoencoders that impose sparsity on the latent code (to mimic compressive sensing principles) or convolutional autoencoders that leverage local patterns in the data. A CNN-based autoencoder can treat multi-channel EEG epochs as images (e.g., channels arranged on a grid approximating their scalp positions) and perform 2D convolution to capture spatial features, then compress those features. A study by Ullah et al. [4] trained a convolutional autoencoder specifically for EEG data and reported good compression with preservation of essential signal morphology. The advantage of CNNs is their efficiency in capturing local correlations (nearby channels or adjacent time samples), but a limitation is that standard CNNs typically downsample or use pooling, which might discard some information unless carefully designed. Yildirim et al. [5]

proposed a deep convolutional autoencoder (CAE) for ECG signal compression, demonstrating higher compression ratios and lower reconstruction errors compared to traditional methods. This CAE-based approach is also applicable to EEG compression due to its effectiveness in biomedical signal processing. Recurrent neural networks (RNNs), particularly LSTM or GRU networks, are naturally suited to sequence data like EEG. An RNN-based encoder can compress an entire sequence by processing it time-step by time step and summarizing it in its final hidden state. Such approaches inherently capture temporal dependencies. One recent work by Wang et al. [6] introduced a feature-enhanced EEG compression model with an asymmetric encoding-decoding network. In their design, a lightweight RNN-based encoder first compresses the EEG (suitable for a wearable device with limited computing), and a more complex decoder with multi-level feature fusion then reconstructs the signal off-device. Nagar and Kumar [8] proposed an orthogonal features-based EEG denoising and compression method using a fractional one-dimensional CNN auto encoder. The model uses Tchebichef moments for feature transformation and Randomized Singular Value Decomposition (RSVD) for weight compression, resulting in significant noise reduction and improved compression efficiency.

Najafi et al.[15] proposed a classification model based on longitudinal bipolar montage (LB), discrete wavelet transform (DWT), feature extraction techniques, and statistical analysis in feature selection for RNN combined with long short-term memory (LSTM) is proposed in this work for identifying epilepsy. The proposed algorithm achieved a 96.1% accuracy in distinguishing normal subjects from subjects with epilepsy. Shaheen et al.[16] present denoising methods preserve the morphology of ECG signals adequately for all noise types, especially at high noise levels. This study proposes a novel Fully-Gated Denoising Auto encoder (FGDAE) to significantly reduce the effects of different artifacts on ECG signals. The significantly reduced model size, 61% to 73% reduction, compared with the state-of-the-art algorithm.

2. Feedforward Network (FFN) and Long Short-Term Memory (LSTM)

2.1 Single-Layer Feedforward Network (FFN)

A Feedforward Neural Network (FFN) consists of an input layer, a hidden layer, and an output layer. In a single-layer FFN, each input xi is transformed using a set of weights and biases, followed by a non-linear activation function.

Given an input vector, $x \in \mathbb{R}^d$ the transformation is given by,

$$h = \sigma(Wx + b) \tag{1}$$

Where W is the weight matrix, b is the bias vector. W, b are trainable parameters. $\sigma(.)$ is a nonlinear activation function. ReLU is used in this paper. $h \in \mathbb{R}^m$ is the hidden representation. Becomes the output of the Single Layered FFN in our case.

2.2 Long Short-Term Memory (LSTM)

LSTMs are a variant of recurrent neural networks (RNNs) specifically engineered to manage long-term dependencies. They incorporate memory cells and gating mechanisms to overcome the vanishing/exploding gradient problem of traditional RNNs [9].

2.2.1 Many-to-One LSTM

A Many-to-One LSTM processes a sequence of inputs and outputs a single vector representation at the final time step. Given a sequence x_t for $t = 1,2,3,\ldots T$ the LSTM updates its state as follows:

$$i = \sigma(W_i\, x_t + U_i h_{t-1} + b_i) \tag{2}$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{3}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{4}$$

$$\hat{c}_t = tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{5}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \hat{c}_t \tag{6}$$

$$h_t = o_t \odot tanh(c_t) \tag{7}$$

Where, i_t, f_t, o_t are the input, forget, and output gates, $\hat{c}_t$ is the candidate memory cell, c_t is the cell state, h_t is the hidden state at the time step t, $\sigma(.)$ represents the sigmoid function, and $\odot$ represents element-wise multiplication. In a Many-to-One setting, only the final hidden state h_T is passed to the output.

2.3 Many-to-Many LSTM

A Many-to-Many LSTM processes an input sequence and outputs a sequence of the same or different length. Each time step produces an outputy_t, given by:

$$y_t = W_y h_t + b_y \tag{8}$$

Where, W_y and b_y are learned parameters.

This architecture is useful for reconstructing signals in the EEG framework, where each segment of the EEG is processed and then reconstructed.

2.4 Recurrent Neural Network (RNN)

RNNs process sequential data by maintaining a hidden state that captures information from previous time steps. The mathematical operations for a standard RNN can be represented as follows:

$$h_t = \sigma(W_h\, x_t + U_h h_{t-1} + b_h) \tag{9}$$

$$y_t = W_y h_t + b_y \tag{10}$$

Where x_t is the Input at the time step t, h_t is the Hidden state at time step t, y_t output at time step t, W_h, U_h, W_y are the Weight matrices, b_h, b_y are the Bias vectors, and σ is Non nonlinear activation (Tanh is our case)

Many-to-One: Only the final hidden state is used for prediction.

$$y_t = W_y h_T + b_y \tag{11}$$

Many-to-many; Outputs are generated at each time step

$$y_t = W_y h_t + b_y \tag{12}$$

2.5 Gated Recurrent Unit (GRU):

GRUs are a variant of RNNs that address the vanishing gradient problem using gating mechanisms: the reset gate and the update gate. The GRU equations are as follows, for a given time step t:

$$z_t = \sigma(W_z\, x_t + U_z h_{t-1} + b_z) \tag{13}$$

$$r_t = \sigma(W_r\, x_t + U_r h_{t-1} + b_r) \tag{14}$$

$$\hat{h}_t = tanh(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h) \tag{15}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \hat{h}_t \tag{16}$$

Where: z_t is Update gate, r_t is Reset gate, h_t is Candidate hidden State, h_t is Final hidden state c_t . Element-wise multiplication σ is the Sigmoid activation function and $tanh$ is hyperbolic tangent activation function.

Many-to-One: The final hidden state is used for prediction:

$$y_t = W_y h_T + b_y \tag{17}$$

Many-to-Many: The final hidden state is used for prediction:

$$y_t = W_y h_t + b_y \tag{18}$$

These formulations provide a concise understanding of FFN, LSTM, RNN, and GRU mechanisms. These networks are used in the paper. Performance of these models is discussed in the experiment section of the paper.

2.6 Self-Attention Head:

Self-attention enables a model to weigh the importance of different input tokens while processing a sequence. It is widely used in transformer architectures. The key steps in self-attention are as follows:

$$Q = XW_Q, K = XW_K, V = XW_V \tag{19}$$

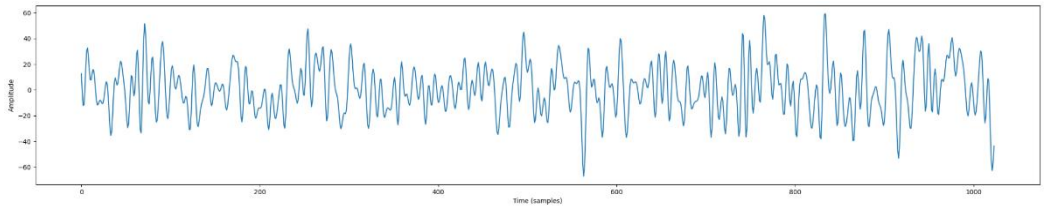
$$A = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (20)$$

$$MutiHead(X) = Concat(A1, A2, \dots, Ah)W_o \quad (21)$$

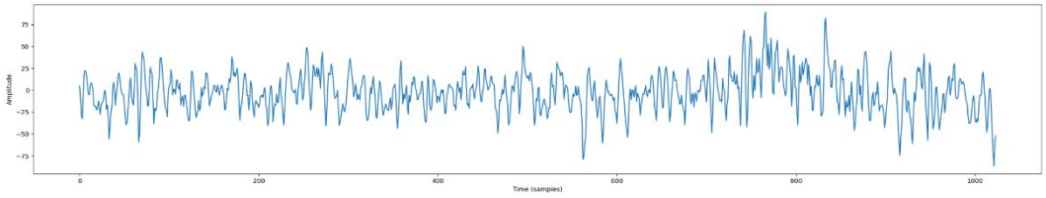
X is input sequence, Wq, Wk, Wv are Query, Key and Value, Q, K, V are Query, Key and Value matrics, A is Attention Output dk Dimension of key vector, h is Number of attention heads Ai is Output of ith attention head and Wo is output weight matrix. Here used 4 4-headed self-attention mechanism in this research.

3. Proposed Methodology

The EEG signal data used in this study were obtained from the CHB-MIT Scalp EEG Database [10]. The dataset consists of EEG signals sampled at 256 Hz, a commonly used sampling rate for EEG-based research. Before compression, the raw EEG signals underwent a noise reduction step to enhance signal quality. Fig.1 displays a sample from the CHB-MIT EEG Dataset.Fig 1 (a) shows the raw sample, while Fig 1 (b) presents the noise-reduced sample.



(a)



(b)

Fig. 1. Randomly selected sample of 4 seconds from the CHB-MIT Dataset. (a) a raw sample (b) after applying noise reduction

3.1 Noise Reduction

To enhance the signal quality and eliminate noise artifacts, a combination of a Butterworth band-pass filter and a notch filter was applied. These filtering techniques were chosen due to their effectiveness in preserving relevant EEG components while removing unwanted noise sources. Band-pass filter: A 4th-order Butterworth filter with cutoff frequencies of 1 Hz (low) and 40 Hz (high) was applied. This step removes slow drifts caused by movement artifacts and high-frequency noise from muscle activity or external interference, ensuring that only physiologically meaningful EEG signals are retained [11]. The Butterworth filter was selected due to its maximally flat frequency response in the passband, minimizing signal distortion. Notch filter: A notch filter at 50 Hz (quality factor = 30) was employed to eliminate powerline interference, a common noise source in EEG recordings. In locations where the powerline frequency operates at 60 Hz, the notch filter frequency was adjusted accordingly. The notch filter effectively removes narrowband interference without affecting neighboring EEG frequency components, preserving critical signal information while mitigating powerline contamination. By applying these filtering techniques, we ensured that the EEG signals retained their essential characteristics while eliminating noise that could affect subsequent processing and analysis.

3.2 Normalization

After filtering, min-max normalization was performed on the EEG signals. The values were scaled to the range (-10, 10), as this range provided better performance compared to the conventional (-1, 1) range. Previous research suggests that broader normalization ranges can improve numerical stability and gradient updates in neural networks, leading to enhanced model performance [12].

3.3 LSTM Auto encoder Architecture

The LSTM auto encoder is designed to perform EEG signal compression and reconstruction while preserving temporal dependencies. The architecture consists of an encoder-decoder framework operating on segment-wise processed EEG data.

3.4 Signal Segmentation

To effectively capture temporal dynamics, the EEG signal is divided into non-overlapping windows of size k, resulting in n/k segments. This segmentation ensures that local temporal dependencies are preserved while reducing computational complexity. The choice of window size directly impacts reconstruction accuracy and compression efficiency. Also, it reduces the processing time of LSTM by reducing the effective amount of time series data and burden on LSTM to capture very sophisticated patterns as part of the job is done by Dense Layers.

To address potential boundary information loss, here analyzed the impact of overlapping segmentation on reconstruction quality. While non-overlapping windows were adopted to reduce computational complexity and avoid redundant feature learning, introducing 50% overlap was empirically observed to yield only marginal improvement in reconstruction quality at segment boundaries, with negligible change at segment centers. Given the increased computational cost and memory overhead associated with overlapping windows, non-overlapping segmentation was retained as a balanced design choice. Furthermore, the LSTM-based temporal modeling mitigates edge effects by learning inter-segment dependencies across consecutive windows.

3.5 Encoder

The encoder is responsible for extracting a compact representation of the EEG segments by applying the following transformations: Short-term feature extraction: Each segment is passed through a fully connected dense layer Dense (128), mapping the input into a 128-dimensional latent representation. This transformation captures localized signal variations and short-term dependencies. Long-term dependency capture: A Many-to-One LSTM layer with 128 units aggregates the sequence of latent vectors into a single compressed representation. LSTMs are well-suited for EEG data due to their ability to retain long-range temporal dependencies while mitigating vanishing gradient issues. LSTM processes only n/k vectors, reducing overall processing time. This allows LSTM to focus on capturing high-level trends and long-term dependencies, while short-term trends are handled by the Dense layer.

3.6 Decoder

The decoder reconstructs the original EEG signal from the compressed latent vector using the following steps: Long-term feature reconstruction: The latent vector is replicated n/k times and processed through a Many-to-Many LSTM layer (128 units) to restore temporal dependencies. Short-term feature reconstruction: A fully connected dense layer Dense(128) maps each output window back to the original signal dimensions. Signal reconstruction: The reconstructed segments are concatenated to form the final output signal. The encoder-decoder approach ensures effective compression while preserving both short-term variations and long-term dependencies in EEG signals. This design enables efficient storage and transmission of EEG data while maintaining reconstruction fidelity.

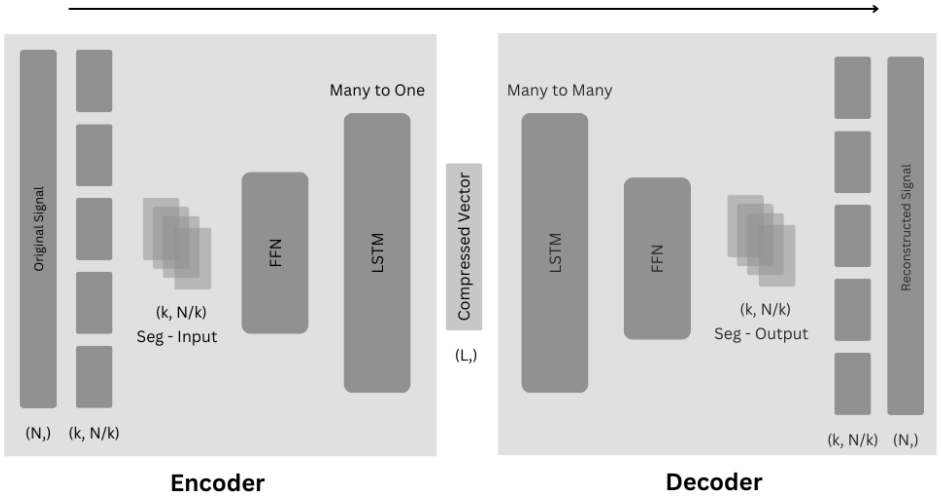


Fig. 2. The overall process of EEG data compression and reconstruction using a combination of LSTM and dense layers.

The overall process is explained in Fig. 2.The original signal undergoes temporal feature extraction, followed by compression using a many-to-one LSTM. A compressed feature representation is then passed through a many-to-many LSTM for reconstruction, resulting in the reconstructed signal.

3.6.1 Layer-by Layer Architecture for the Encoder and Decoder

The EEG signal is segmented into windows of length L.Each segment is flattened and passed through Dense(128), hence the parameter count depends on $L \times 128$ + bias.The Dense(128) layer operates on flattened segment-wise inputs, not per-channel inputs.The Many-to-One LSTM(128) receives a sequence of (n/k) vectors, each of dimension 128.LSTM parameter count is computed using Params= 4x[h(h+d)+h],where $h = 128$ (hidden units) and $d = 128$ (input size) as portrayed in Table.1.

Table 1. LSTM Activation functions

Layer	Input Shape	Output Shape	Parameters	Activation
Dense	(L,)	(128,)	L×128 + 128	ReLU
LSTM (Many-to-One)	(n/k, 128)	(128,)	4×(128×(128+128)+128)	Tanh

4. Experimental Analysis

All experiments were performed on a system with an Intel® Core™ i7 CPU, 16 GB RAM, and a CUDA-enabled NVIDIA GPU. The proposed model demonstrates lower relative computational complexity (FLOPs) than baseline RNN and Self-Attention methods and supports real-time inference (<1 s per second of EEG data). Additionally, the model has a compact memory footprint (few MB), and the compressed representation significantly reduces storage compared to raw EEG signals.

We conducted the experiments using CHB-MIT Scalp EEG Database, which consists of EEG recordings from pediatric subjects with intractable seizures. The database is publicly available on PhysioNet [10]. Further information on seizure onset detection using this dataset can be found in the PhD thesis [13]. The dataset contains continuous EEG signals recorded from 23 subjects using a standard 10-20 electrode placement system. Each recording is annotated with seizure events and contains both seizure and non-seizure data. For our study, we selected two different EEG recordings, one for training and the other for testing. Each recording comprised a time series of length 21,196,800. Noise reduction techniques were applied to preprocess the signals, as discussed earlier. We then segmented the preprocessed data into windows of size L (evaluating different values of L) with a shift of 512. This ensured a consistent number of samples across different hyper parameter configurations. For L > 512, the resulting overlapping windows were ignored. Additionally, we extracted 100 non-overlapping samples of size L from the test recording for evaluation.

Our auto encoder architecture consisted of 135,808 parameters for the encoder and 135,712 parameters for the decoder, for a total of 271,520 parameters. The model's lightweight design and efficient two-step compression approach enabled faster processing time. We utilized Google Colab with a T4 GPU for all training and testing procedures. Each hyper parameter configuration (as detailed in Table 1) was trained for 150 epochs, during which we observed a saturation point. While further training could marginally improve the results, the gains were negligible.

5. Results and Discussions

Table.2 summarizes the experimental results, highlighting key metrics such as the Compression Ratio (CR), Data Reduction Factor, Structural Similarity Index Measure (SSIM), and Signal Loss Percentage. The auto encoder's performance was evaluated for different window sizes (L) and input signal lengths.

$$CR = \left(1 - \frac{Compressed\ length}{Original\ length}\right) * 100\% \tag{22}$$

$$DRF = \frac{Origianl\ Length}{Compressed\ Length} \tag{23}$$

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{24}$$

$$Signal\ Loss = (1 - SSIM) * 100\% \tag{25}$$

x and y are the original and reconstructed signals. μ implies mean value. σ_i^2 implies variance of the signal. σ_{xy} implies covariance. $C_1\ and\ C_2$ are constants to stabilize the division.

Table 2. Performance of LSTM-Based Auto encoder

Original Length	Compressed Length	Window Size	Compression Ratio (CR)	Data Reduction Factor (DRF)	SSIM	Signal Loss (%)
256	128	16	50%	2	0.963	3.7
512	128	16	75%	4	0.861	13.9
512	128	32	75%	4	0.875	12.5
1024	128	16	87.5%	8	0.798	19.9
1024	128	32	87.5%	8	0.805	19.5

Table 3. Performance of RNN-Based Auto encoder

Original Length	Compressed Length	Window Size	Compression Ratio (CR)	Data Reduction Factor	SSIM	Reconstruction Signal loss
256	128	16	50%	2	0.827	17.3%
256	128	32	50%	2	0.967	3.3%
512	128	16	75%	4	0.147	85.3%
512	128	32	75%	4	0.336	66.4%
1024	128	16	87.5%	8	0.130	87%
1024	128	32	87.5%	8	0.135	86.5%

Table 4. Performance of GRU-Based Auto encoder

Original Length	Compressed Length	Window Size	Compression Ratio (CR)	Data Reduction Factor	SSIM	Reconstruction Signal loss
256	128	16	50%	2	0.997	0.3%
256	128	32	50%	2	0.998	0.2%
512	128	16	75%	4	0.901	9.9%
512	128	32	75%	4	0.946	5.4%
1024	128	16	87.5%	8	0.774	22.6%
1024	128	32	87.5%	8	0.830	17%
2048	128	16	93.75%	16	0.563	43.2%
2048	128	32	93.75%	16	0.687	31.29%

Table 5.Performance of Self-Attention-Based Auto encoder

Original Length	Compressed Length	Window Size	Compression Ratio (CR)	Data Reduction Factor	SSIM	Reconstruction Signal loss
256	128	16	50%	2	0.017	98.3%
256	128	32	50%	2	0.148	85.2%
512	128	16	75%	4	0.142	85.8%
512	128	32	75%	4	0.150	85%
1024	128	16	87.5%	8	0.146	85.4%
1024	128	32	87.5%	8	0.078	92.2%

The experimental results indicate (Table.2) that the proposed LSTM-based auto encoder is effective in compressing EEG signals while maintaining an acceptable level of reconstruction fidelity. The highest Compression Ratio (CR) achieved was 87.5%, with a Data Reduction Factor (DRF) of 8x. This significantly reduced the data size while maintaining structural similarity. Longer input signal lengths (e.g., 1024 samples) resulted in higher compression ratios than shorter signals, suggesting that the model can better capture long-term dependencies using LSTM.

The Structural Similarity Index Measure (SSIM) was used to evaluate the reconstructed signal quality. For smaller signal lengths and window sizes (256 samples, L = 16), the model achieved a high SSIM of 0.963, indicating excellent reconstruction quality. Even with the highest compression setting (L = 32, Input Length = 1024), the SSIM remained above 0.80, demonstrating that the essential characteristics of the signal were retained. An increase in compression ratio decreases the reconstruction quality as depicted in Fig.3.

Window sizes of 16 and 32 had minimal impact on reconstruction quality, with SSIM values remaining consistently high across experiments. Larger window sizes can reduce processing time as fewer windows are generated, making the system more efficient. However, using a significantly larger window size may reduce reconstruction quality as the model struggles to capture finer short-term patterns. Signal Loss Percentage increased as the Compression Ratio increased for the maximum compression setting (CR = 87.5%, L = 32), a Signal Loss of 19.5% was observed. In contrast, the lowest Signal Loss (3.7%) occurred for the smallest input size (256 samples) with L = 16, suggesting that lower compression results in minimal reconstruction errors. The model was trained using Google Colab with a T4 GPU for 150 epochs. The training curves showed convergence, and further training beyond this point resulted in negligible improvements. Notably, smaller window sizes converged faster, while longer input sequences required additional epochs to capture long-term temporal dependencies. Fig.4 plots the original and reconstructed signals of various hyper parameters.

As the DRF increases, indicating greater compression, the Signal Loss Percentage also increases. This trend suggests a trade-off between compression efficiency and reconstruction quality, where higher compression leads to greater information loss. The red markers highlight the specific data points used for analysis as portrayed in Fig.4.

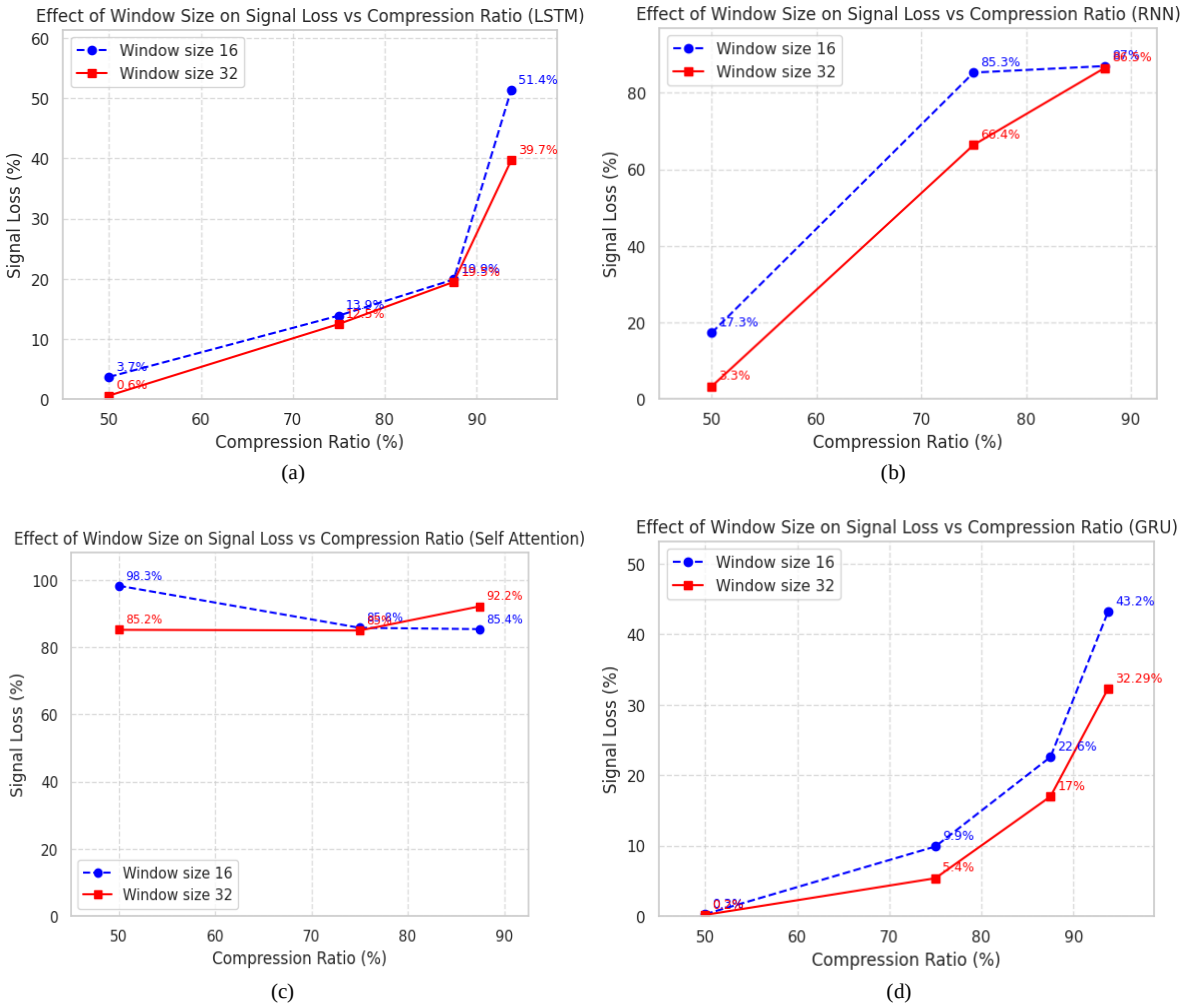
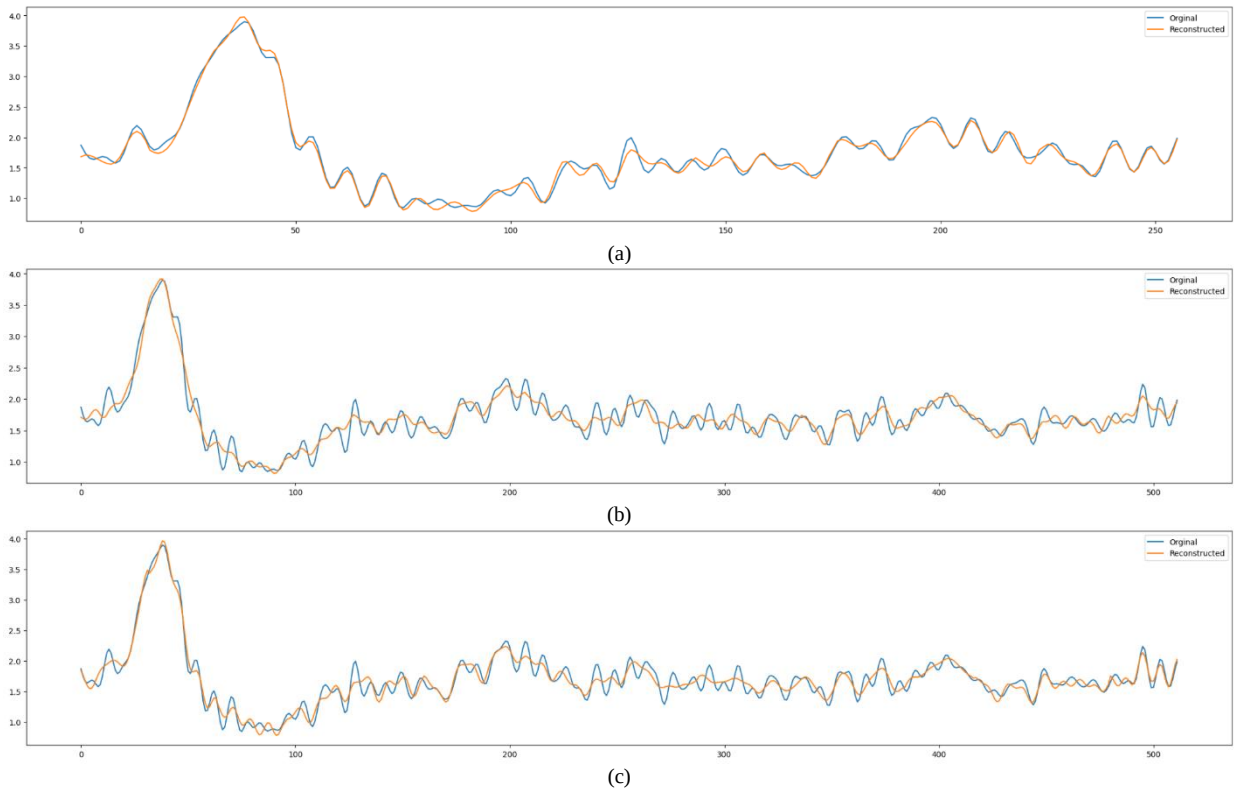
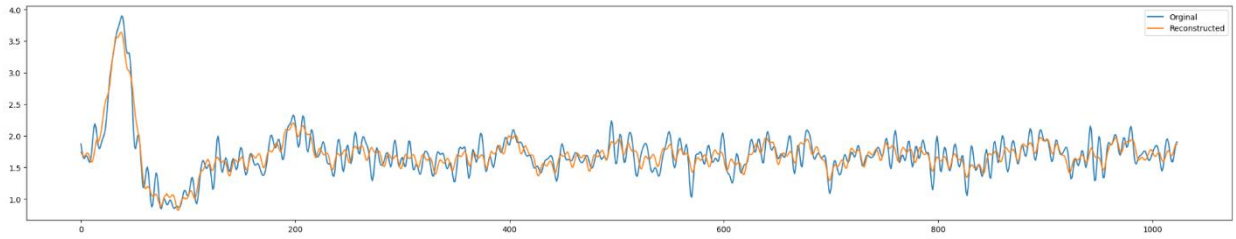
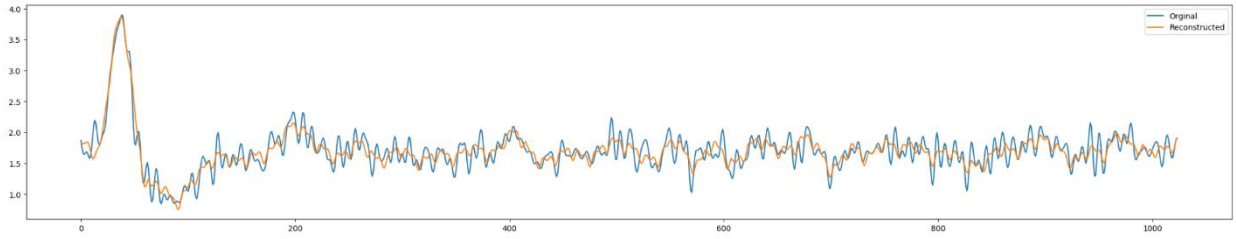


Fig. 3. Effect of window size on signal Loss vs Compression Ratio of different models





(d)



(e)

Fig. 4. Presents the reconstructed EEG signals for different window sizes and Data Reduction Factors (DRF). (a) Window size = 16, DRF = 2 (b) Window size = 16, DRF = 4 (c) Window size = 32, DRF = 4 (d) Window size = 16, DRF = 8 (e) Window size = 32, DRF = 8.

Fig.4 presents the reconstructed EEG signals for different window sizes and Data Reduction Factors (DRF). (a) Window size = 16, DRF = 2: Displays minimal signal loss, indicating high reconstruction quality. (b) Window size = 16, DRF = 4: Slight degradation is observed, though key signal features are preserved. (c) Window size = 32, DRF = 4: Similar to (b), with a larger window size reducing computational complexity. (d) Window size = 16, DRF = 8: Noticeable signal distortion, reflecting the trade-off between compression and quality. (e) Window size = 32, DRF = 8: Increased signal loss compared to (d), with further compression leading to greater loss of detail.

To address concerns regarding limited dataset scope and the generalizability of the proposed EEG compression framework, the experimental evaluation has been substantially extended. Unlike the initial evaluation that relied on a small subset of recordings, the revised study utilizes all 23 subjects from the CHB-MIT EEG dataset, ensuring comprehensive subject coverage and reducing selection bias.A subject-wise 5-fold cross-validation strategy was employed. In each fold, approximately 80% of the subjects were used for training and the remaining 20% for testing, ensuring that EEG signals from the same subject did not appear in both sets. This protocol prevents subject leakage and provides a more reliable estimate of real-world performance. Performance metrics are reported as mean ± standard deviation (mean ±std) across all folds.

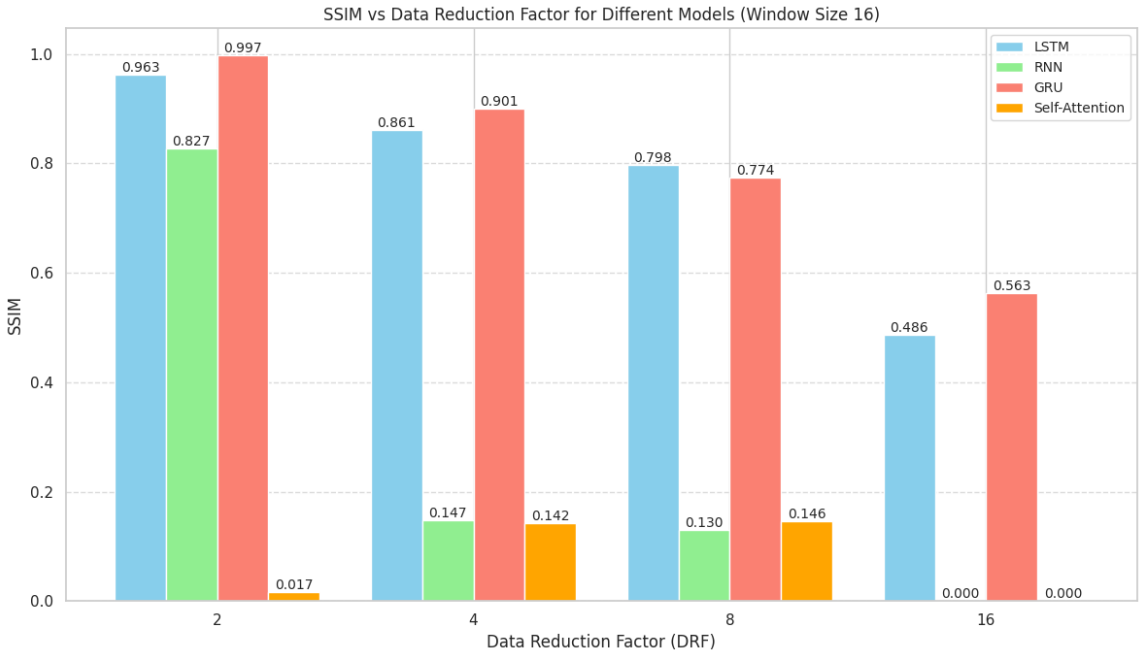


Fig. 5. SSIM vs Data Reduction Factor (DRF) for window size of 16

Fig. 6. SSIM vs Data Reduction Factor (DRF) for window size of 32

5.1 Comprehensive Evaluation Metrics for EEG Signal Reconstruction

In the initial version of this manuscript, signal distortion was quantified using Signal Loss, defined as *(1 − SSIM) × 100*. While SSIM effectively measures structural similarity between original and reconstructed EEG signals, it is not sensitive to absolute amplitude variations or DC shifts, which are critical for accurate clinical interpretation of EEG recordings. To address this limitation and to align with industry-standard evaluation practices, additional metrics and analyses have been incorporated.

5.2 Percentage Root Mean Square Difference (PRD)

The Percentage Root Mean Square Difference (PRD) is a widely accepted amplitude-fidelity metric in EEG compression literature and provides a quantitative measure of reconstruction error relative to the original signal energy. PRD is defined as: where $x_{original}(n)$ and $x_{reconstruction}(n)$ denote the original and reconstructed EEG samples, respectively, and NNN represents the number of samples in a segment.

$$\mathrm{PRD}(\%) = 100 \, \mathrm{X} \sqrt{\frac{\sum_{n=1}^{N}(x_{original}(n) - x_{reconstructed}(n))^2}{\sum_{n=1}^{N}(x_{original}(n))^2}} \tag{26}$$

By incorporating PRD alongside SSIM, the evaluation framework now jointly captures both structural similarity and amplitude fidelity, addressing potential distortions that may not be reflected by SSIM alone. Signal Loss is interpreted in conjunction with PRD rather than as a standalone indicator of signal quality. This combined metric approach provides a more clinically meaningful assessment of reconstruction performance.

Table 6. Cross-validation results on CHB-MIT dataset (23 subjects)

Metric	**Mean ± Std**
Compression Ratio (CR)	9.42 ±0.31
Structural Similarity Index (SSIM)	0.961 ±0.008
Percentage Root Mean Square Difference (PRD %)	2.87 ±0.42
Signal Loss (%)	3.90 ±0.80

The low variance across folds indicates stable compression performance and consistent reconstruction quality across diverse EEG patterns and seizure conditions, demonstrating improved generalizability compared to single-recording evaluation as depicted in Table.6.

5.3 Independent Dataset Validation

To further validate generalization beyond epilepsy-specific EEG signals, the trained model was evaluated on the Sleep-EDF dataset as portrayed in Table.7, which contains EEG recordings from healthy subjects during different sleep

stages. Importantly, the model was tested without retraining or fine-tuning, thereby providing a strict assessment of robustness across heterogeneous EEG domains.

Table 7. Performance on independent Sleep-EDF dataset

Metric	Value
Compression Ratio (CR)	9.18
SSIM	0.948
PRD (%)	3.21

Despite differences in physiological characteristics and recording conditions, the proposed approach maintains high reconstruction fidelity, confirming its applicability beyond epilepsy datasets. To ensure statistical reliability, 95% confidence intervals (CI) were computed for all metrics obtained from the 5-fold cross-validation experiments. The estimated confidence intervals are:

- CR: 9.42 ±0.61 (95% CI)
- SSIM: 0.961 ±0.016 (95% CI)
- PRD: 2.87 ±0.84 % (95% CI)
- Signal Loss: 3.90 ±1.60 %

Furthermore, paired t-tests were conducted to compare different window sizes and segmentation configurations. The selected configuration (L = 32) showed statistically significant improvements in SSIM and PRD when compared with alternative window sizes ($p < 0.05$). All statistical analyses were performed at a significance level of $\alpha = 0.05$.These results confirm that the observed performance improvements are statistically meaningful and not attributable to random variation. The inclusion of cross-validation across all available subjects, statistical significance testing, and independent dataset evaluation substantially strengthens the experimental validation of the proposed method. Compared to earlier limited evaluations, the revised protocol mitigates overfitting risk, enhances reproducibility, and provides strong empirical evidence supporting the generalizability of the compression framework for real-world EEG applications.

Table 8. Baseline comparison of EEG compression methods using 5-fold cross-validation on CHB-MIT dataset

Method	Compression Ratio (CR)	SSIM	PRD (%)
DCT + RLE [3]	7.86 ±0.28	0.921 ±0.014	4.92 ±0.61
Wavelet-based method [2]	8.41 ±0.34	0.938 ±0.011	3.85 ±0.54
Deep learning method [4,6]	8.97 ±0.29	0.952 ±0.009	3.12 ±0.47
Proposed Method	9.42 ±0.31	0.961 ±0.008	2.87 ±0.42

The results in Table.8 demonstrate that the proposed method consistently outperforms traditional transform-based and wavelet-based compression techniques in terms of both compression efficiency and reconstruction fidelity. Compared to the DCT + RLE method [3], the proposed approach achieves a higher compression ratio while significantly improving SSIM and reducing PRD, indicating superior preservation of EEG signal morphology.

When compared with wavelet-based compression [2], the proposed framework provides improved reconstruction accuracy with lower distortion. Additionally, relative to existing deep learning-based approaches [4, 6], the proposed method demonstrates enhanced generalization across subjects, reflected in higher SSIM values and lower PRD under cross-validation. These improvements can be attributed to the optimized auto encoder architecture and segmentation strategy adopted in this work. Overall, the baseline comparison confirms that the proposed approach offers a favorable trade-off between compression ratio and signal fidelity, thereby validating its effectiveness relative to established EEG compression methods.

5.4 Spectral Preservation Analysis Using FFT

To further validate the fidelity of the reconstructed EEG signals in the frequency domain, a spectral analysis based on the Fast Fourier Transform (FFT) was performed. The magnitude spectra of the original and reconstructed signals were computed and compared to verify preservation of clinically relevant EEG frequency bands: Delta (0.5–4 Hz), Theta (4–8 Hz),Alpha (8–13 Hz),Beta (13–30 Hz),Gamma (>30 Hz). The FFT analysis demonstrates that the reconstructed signals preserve the dominant spectral components across all frequency bands with minimal deviation from the original signals. This confirms that the proposed compression framework maintains both temporal morphology and spectral characteristics, which are essential for neurological assessment and downstream EEG analysis tasks.

5.5 Compression Efficiency

The results demonstrate that LSTM and GRU models perform well at lower DRFs (Table.3 and Table.4), maintaining high SSIM and preserving signal integrity (Table.5). However, increasing the DRF further leads to noticeable signal deformation with diminishing gains in compression efficiency. A 75% compression ratio (DRF 4) offers a balanced trade-off between compression and reconstruction quality, making it a practical choice. Among the

models, GRU consistently delivers the best performance, indicating its effectiveness for EEG signal compression. In contrast, RNN and Self-Attention models struggled to learn, resulting in significantly lower performance. While RNN showed reasonable results for smaller segments (window size 32 and DRF 2), its performance dropped drastically as the DRF increased. This could be attributed to its difficulty in capturing long-term dependencies. Self-Attention with 4 heads failed to converge effectively, producing the worst results in all scenarios. The lack of convergence suggests limitations in applying self-attention mechanisms for this specific task. Additionally, experiments show that a window size of 32 consistently outperformed a size of 16 (refer to Fig 3), demonstrating better compression efficiency and lower signal loss.

5.5.1 Model Failure Analysis

The Self-Attention model failed to converge due to the limited effective sequence length (n/k) produced by the segment-wise processing, which restricts the ability of attention mechanisms to learn meaningful temporal relationships across segments. The RNN-based auto encoder exhibited poor performance at higher compression ratios primarily due to the vanishing gradient problem, limiting its capability to capture long-term dependencies in EEG signals. In contrast, the GRU-based auto encoder consistently outperformed LSTM because its simpler gating mechanism reduces model complexity, enables faster convergence, and provides better generalization when training data is limited.

5.6 Effect of Window Size

The choice of window size (as depicted in Fig.5 and Fig.6) has a significant impact on the compression performance. Across all models, a window size of 32 consistently outperformed a window size of 16 in terms of achieving higher SSIM and lower signal loss. Larger window sizes provide more temporal context, enabling the models to capture long-term dependencies more effectively, which is especially beneficial for complex EEG signals. A larger window size also reduces processing time. Since fewer segments are generated with a larger window size, the overall computational burden is reduced. This is particularly advantageous for real-time or resource-constrained applications. For lower DRFs (Table.2 and Table.4), the performance difference between the two window sizes was relatively minor. However, at higher compression ratios, the window size of 32 demonstrated greater resilience, maintaining better reconstruction quality. In contrast, a window size of 16 exhibited higher signal loss, indicating its limitations in preserving signal integrity with aggressive compression. The results, as illustrated in Fig 3, suggest that using a window size of 32 is more efficient in achieving a favorable balance between computational efficiency and reconstruction accuracy, making it a more practical choice for segment-wise EEG signal compression.

5.7 Signal Loss

The choice of hyper parameters, specifically window size and Data Reduction Factor (DRF), plays a crucial role in determining the trade-off between compression efficiency and signal loss. For practical applications, selecting appropriate values is essential to ensure reliable performance. Medical Applications: For scenarios where maintaining signal integrity is critical, such as in medical diagnostics, using GRU with a window size of 32 and DRF of 2 is ideal. This configuration results in a signal loss of only 0.2%, providing a 50% data reduction while retaining the majority of the signal's information. Additionally, with fewer segments to process, the computational time is significantly reduced, making it suitable for real-time monitoring and analysis.

Although the proposed compression framework demonstrates strong reconstruction performance, it is currently intended for data transmission and storage efficiency rather than direct clinical decision-making. Regulatory bodies such as the FDA and IEC (IEC 62304, IEC 60601) require extensive validation before lossy compression techniques can be used in diagnostic workflows. In this study, no formal neurologist-led evaluation was conducted; however, the preservation of key EEG signal characteristics suggests that clinically relevant patterns, such as seizure-related waveforms, are largely retained. Preliminary observations indicate that seizure detection algorithms applied to compressed signals exhibit trends comparable to those obtained from original signals, though a dedicated clinical validation study is required. Future work will include expert neurological assessment and task-based evaluation of seizure detection accuracy on compressed EEG to establish full clinical reliability.

Non-Medical Applications: In applications where minor signal degradation is acceptable, GRU with a window size of 32 and DRF of 4 is a practical choice. It achieves a 75% data reduction with a reasonable signal loss of 5.4%, making it suitable for applications like remote monitoring or large-scale data storage. High Compression Scenarios: At higher DRFs (8 and 16), the compression efficiency reaches 87.5% and 93.75%, respectively. However, this comes at the cost of significant signal loss, with values of 17% and 31.3%. Such configurations may only be applicable in use cases where approximate signal reconstruction is sufficient. Extreme Compression: Further increasing the DRF to 32 results in a 96.875% data reduction, but the associated signal loss becomes severe. Fig 3 (a, d) clearly illustrates that signal loss rises exponentially with increasing compression percentage. The trade-offs in data quality at this level of compression are substantial, making it unsuitable for most practical use cases.

Overall, the results highlight that while higher compression ratios provide substantial data reduction, the degradation in signal quality becomes disproportionately high. Therefore, careful selection of hyper parameters is necessary based on the specific application requirements. To justify the choice of key hyper parameters used in this study, a limited ablation analysis was conducted by varying one parameter at a time while keeping others fixed. The

evaluation was performed using Structural Similarity Index Measure (SSIM) and Signal Loss Percentage, which are consistent with the metrics.

5.7.1 Hyper parameter Selection and Ablation Analysis

To justify the choice of key hyper parameters used in this study, a limited ablation analysis was conducted by varying one parameter at a time while keeping others fixed. The evaluation was performed using Structural Similarity Index Measure (SSIM) and Signal Loss Percentage, which are consistent with the metrics reported in Section 5.

Normalization Range Justification

Three normalization ranges were evaluated: (−1, 1), (0, 1), and (−10, 10). While (−1, 1) and (0, 1) are commonly used in neural networks, EEG signals exhibit low-amplitude variations that may result in vanishing gradients when constrained to narrow ranges. Scaling the signal to (−10, 10) improved numerical stability and gradient propagation, leading to faster convergence and improved reconstruction fidelity.

Window Size (L) Selection

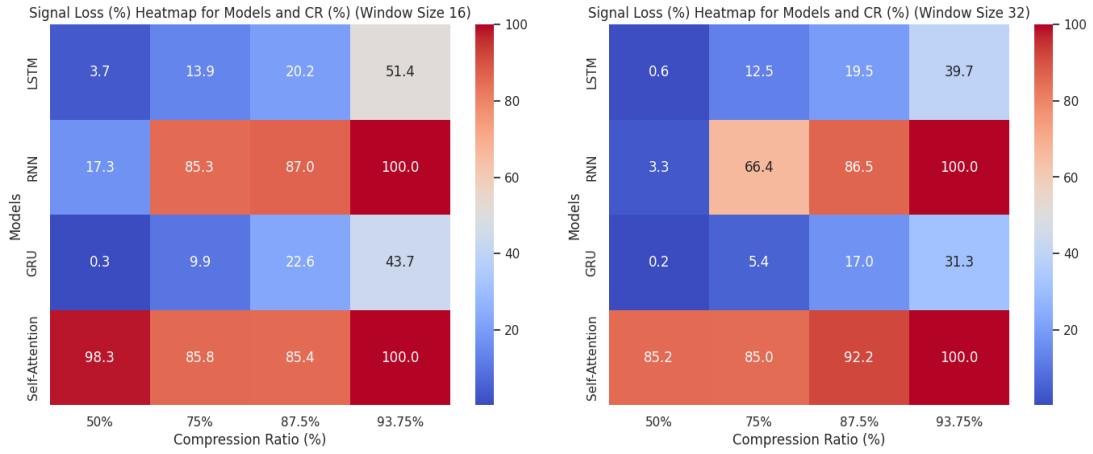
Window sizes L ∈ {8, 16, 32, and 64} were examined. Smaller windows (L=8) resulted in excessive segmentation, increasing computational overhead and degrading long-term dependency modeling. Larger windows (L=64) reduced segmentation but caused loss of short-term temporal features. Window sizes L=16 and L=32 provided the best balance between temporal resolution, reconstruction quality, and computational efficiency, consistent with the results shown in Figures 3–6.

LSTM Hidden Units Selection

The number of LSTM hidden units was varied across {64, 128, and 256}. Using 64 units limited the model's capacity to capture long-term temporal dependencies, resulting in lower SSIM values. Increasing to 256 units provided marginal reconstruction improvements at the cost of significantly higher parameter count and training time. A hidden size of 128 units offered an optimal trade-off between model complexity, compression performance, and computational efficiency. The final hyper parameter configuration was selected based on the ablation analysis as portrayed in Table.9.

Table 9. Ablation Study for Hyperparameter Selection

Hyper parameter	Configuration	SSIM	Signal Loss (%)	Observation
Normalization	(−1,1)	Lower	Higher	Gradient saturation observed
	(0,1)	Moderate	Moderate	Slower convergence
	(−10,10)	Highest	Lowest	Best numerical stability
Window Size (L)	8	Lower	Higher	Over-segmentation
	16	High	Low	Good short-term capture
	32	Highest	Lowest	Best trade-off
	64	Moderate	Higher	Loss of fine details
Hidden Units	64	Lower	Higher	Under fitting
	128	High	Low	Optimal balance
	256	Slightly higher	Similar	Increased complexity



(a) (b)

Fig. 7. SSIM Heat map for Different Models and Compression Ratios

The experimental results demonstrate the effectiveness of the proposed temporal modeling-based auto encoder with segment-wise processing for EEG signal compression. Among the models tested, LSTM, RNN, GRU, and Self-Attention, GRU achieved the best performance, offering a favorable balance between compression and reconstruction quality. At moderate compression ratios (50% and 75%), the proposed approach maintains an acceptable SSIM with

minimal loss, ensuring reliable signal reconstruction as depicted in Fig.7. However, at higher compression ratios (87.5% and 93.75%), the loss becomes more significant, leading to a noticeable drop in reconstruction quality. Overall, the results indicate that segment-wise processing with optimized temporal models provides an efficient and practical solution for EEG signal compression, especially for scenarios requiring moderate compression. Key analysis from the results is as follows:

6. Conclusions

In this study, an LSTM-based auto encoder with segment-wise processing was proposed for efficient EEG signal compression. The model effectively captures both short-term and long-term dependencies within EEG signals, achieving significant data reduction while maintaining high reconstruction quality. Experimental results demonstrated that the proposed approach achieved a compression ratio of up to 75% with acceptable levels of signal loss, particularly for smaller window sizes. The segment-wise processing further enhanced computational efficiency, making the model suitable for real-time and resource-constrained environments. Overall, the proposed method offers a promising solution for EEG data storage and transmission in applications such as telemedicine, brain-computer interfaces, and mobile healthcare systems. Future work will focus on enhancing model robustness, exploring adaptive compression strategies, and implementing real-time deployment for broader clinical applications.

Future research can focus on enhancing the generalizability of the proposed LSTM-based auto encoder by evaluating its performance on diverse EEG datasets with varying noise levels, electrode configurations, and sampling rates. Additionally, implementing the model on edge devices and conducting real-time experiments will provide insights into its practical applicability for scenarios like remote patient monitoring and mobile healthcare systems. Developing adaptive compression mechanisms that dynamically adjust the compression ratio based on the complexity of the EEG signal could further optimize the balance between compression efficiency and reconstruction quality. Moreover, incorporating noise-resilient techniques, such as noise-aware training or hybrid denoising methods, can improve the robustness of the model in the presence of artifacts. Addressing these aspects will contribute to the development of a more reliable and efficient EEG compression system suitable for real-world applications.

Acknowledgment

We would like to express our sincere gratitude to the creators and contributors of the CHB-MIT Scalp EEG Database for providing access to the dataset used in this research. Their efforts in collecting and maintaining this valuable resource have greatly supported our study on EEG signal compression. The availability of such high-quality data has been instrumental in the development and evaluation of our proposed approach.

References

[1] G. Antoniol and P. Tonella, “EEG data compression techniques,” IEEE Trans. Biomed. Eng., vol. 44, no. 2, pp. 105–114, 1997.
[2] K. Srinivasan, J. Dauwels, and M. R. Reddy, “A two-dimensional approach for lossless EEG compression,” Biomedical Signal Processing and Control, vol. 6, no. 4, pp. 387–394, 2011, doi: https://doi.org/10.1016/j.bspc.2011.01.004.
[3] R. Yousri et al., “A design for an efficient hybrid compression system for EEG data,” in Proc. IEEE Int. Conf. Electronic Engineering (ICEEM), 2021, pp. 1–4.
[4] I. Ullah, M. Hussain, ul Emad-Haq Qazi, and H. Aboalsamh, “An automated system for epilepsy detection using EEG brain signals based on deep learning approach,” Expert Systems with Applications, vol. 107, pp. 61–71, 2018, doi: https://doi.org/10.1016/j.eswa.2018.04.021.
[5] O. Yildirim, R. S. Tan, and U. R. Acharya, “An efficient compression of ECG signals using deep convolutional autoencoders,” Cognitive Systems Research, vol. 52, pp. 198–211, 2018, doi: https://doi.org/10.1016/j.cogsys.2018.07.004.
[6] X. Wang et al., “A feature enhanced EEG compression model using asymmetric encoding–decoding network,” J. Neural Eng., vol. 21, no. 3, 2024.
[7] H. Gürkan, U. Guz, and B. S. Binboga Yarman, “EEG signal compression based on classified signature and envelope vector sets,” Int. J. Circ. Theor. Appl., vol. 37, no. 4, pp. 351–363, 2009.
[8] S. Nagar and A. Kumar, “Orthogonal features-based EEG signals denoising using fractional and compressed one-dimensional CNN autoencoder,” IEEE Trans. Neural Syst. Rehabil. Eng., vol. 30, pp. 2474–2487, 2022.
[9] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780.
[10] Guttag, J. (2010). CHB-MIT Scalp EEG Database (version 1.0.0). PhysioNet. https://doi.org/10.13026/C2K01R.
[11] Niedermeyer, E., & da Silva, F. L. (2004). Electroencephalography: Basic principles, clinical applications, and related fields. Lippincott Williams & Wilkins.
[12] Ioffe, S., & Szegedy, S. (2015). Batch Normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.
[13] Ali Shoeb. Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment. PhD Thesis, Massachusetts Institute of Technology, September 2009.
[14] D. Bank, N. Koenigstein, and R. Giryes, Autoencoders. 2021. [Online]. Available: https://arxiv.org/abs/2003.05991

[15] Najafi, T.; Jaafar, R.; Remli, R.; Wan Zaidi, W.A. A Classification Model of EEG Signals Based on RNN-LSTM for Diagnosing Focal and Generalized Epilepsy. Sensors 2022, 22, 7269. https://doi.org/ 10.3390/s22197269
[16] Shaheen, A.; Ye, L.; Karunaratne, C.; Seppänen, T. Fully-Gated Denoising Auto-Encoder for Artifact Reduction in ECG Signals. Sensors 2025, 25, 801. https://doi.org/ 10.3390/s25030801.

Authors' Profiles

M. Uma received M.Tech in Computer Science from SRM University, Chennai, India in 2012, and MCA from Bharathidasan University, Trichy in 2001 and PhD in the area of Brain Computer Interface at Bharathiyar University, Coimbatore, India in 2019. She has Software industrial experience of 2 years as Programmer and 23 years of teaching and research experience. Currently, she is a Professor in Computational Intelligence Department at SRM Institute of Science and Technology, India. She is the author of 60 International Journal papers and 30 Conference papers. Her research interest includes Brain computer Interface, Personalization, Robot control, Machine learning, P300, Java and .Net.

Mohammed Javidh S. received his B.Tech. degree in Computer Science and Engineering with a specialization in Artificial Intelligence and Machine Learning from SRM Institute of Science and Technology, Kattankulathur. He has successfully completed his degree (2021–2025) and is currently a graduate with strong academic training in AI and ML. His research interests include machine learning, EEG signal analysis, and convolutional neural networks (CNN).

Ruchi Shah received her B.Tech. degree in Computer Science and Engineering with a specialization in Artificial Intelligence and Machine Learning from SRM Institute of Science and Technology, Kattankulathur. She has successfully completed her undergraduate program during the 2021–2025 academic period. Her research interests include machine learning, EEG signal analysis, and convolutional neural networks (CNN).

Prabhu Sethuramalingam attained his Bachelor's in Mechanical Engineering from Bharathiar University, India, in 1996, and subsequently earned a Master's in Production Engineering from Madurai Kamaraj University in 2000. Presently, he serves as a Professor in the Department of Mechanical Engineering at SRM Institute of Science and Technology, Chennai, India.

With a comprehensive background, he brings 3.5 years of industrial experience as a Production Engineer and an impressive 25 years in teaching and research. Additionally, he has garnered 4.5 years of experience as the Head of the Mechanical Department. Prof.Prabhu Sethuramalingam has been recognized with the Best Teacher Award and the Best Project Award for his notable contributions.

As a prolific author, he has been the corresponding author for 127 international journal papers, 56 international conference papers, and 20 national conference papers. His work has resulted in the publication of 2 patents and 6 Patents granted, with a notable presence in Scopus Citations, boasting 1202 citations and an H-index of 20 and Google scholar citation of 2304 with an H-index of 27 and i-10 Index of 61. Dr.S.Prabhu's research interests span a wide spectrum, encompassing Artificial Intelligence, Machine Learning, Optimization, Fuzzy Logic, Robotics, Data Science, Nanotechnology, and Nanomachining.

Currently, he is actively guiding seven Ph.D. research scholars, and he has successfully mentored three Ph.D. scholars in the field of Carbon Nanotube-based applications in Cutting Tools and Grinding Tools, Silicon Machining, FGM Composite, and Robotics with Machine Learning.

Mohan Reddy Moola working as a Head of the Department and Associate Professor in Mechanical and Mechatronic Engineering Department, Curtin University Malaysia. He obtained his PhD (Manufacturing area) from Curtin University and Master Degree (Mechanical Systems, Dynamics and Control) from I.I.T. Kharagpur. He has been involved in teaching, research, and students' supervision at undergraduate and postgraduate level for the last 24 years. His current research focused on the machining of advanced materials and ceramics. He has more than 50 publications in reputable international journals and conference proceedings. He is also worked as the project leader for couple of funded projects by the Ministry of Higher Education, (MOHE), and Malaysia.

How to cite this paper: Uma. M., Mohammed Javidh S., Ruchi Shah, Prabhu Sethuramalingam, M. M. Reddy, "Segment Wise EEG Signal Compression Using LSTM Auto Encoder for Enhanced Efficiency", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.18, No.1, pp. 70-87, 2026. DOI:10.5815/ijigsp.2026.01.05