Proceedings of the ASME 2026 45th International
Conference on Ocean, Offshore and Arctic Engineering
OMAE2026
June 7-12, 2026, Tokyo, Japan



**OMAE2026-181966**

# MULTI-HEAD ATTENTION-ENHANCED GATE RECURRENT UNITS FOR MOORING TENSION PREDICTION OF TLP FOWTS

**Maokun Ye**
Computational Marine Hydrodynamic Lab (CMHL),
School of Ocean and Civil Engineering,
Shanghai Jiao Tong University,
Shanghai, China

**Yisheng Yao**
Computational Marine Hydrodynamic Lab (CMHL),
School of Ocean and Civil Engineering,
Shanghai Jiao Tong University,
Shanghai, China

**Shiyuan Zhang**
Offshore Oil Engineering Co. Ltd
Tianjin, China

**Shu Dai**
Shanghai Investigation, Design and Research Institute Co. Ltd.
Shanghai, China

**Decheng Wan***
Computational Marine Hydrodynamic Lab (CMHL),
School of Ocean and Civil Engineering,
Shanghai Jiao Tong University,
Shanghai, China

## ABSTRACT

*Tension-Leg Platform (TLP) floating offshore wind turbines (FOWTs) represent a promising solution for deep-water wind energy, combining cost-effectiveness with excellent station-keeping capabilities. Their vertically taut mooring system provides high global stiffness, shifting natural frequencies above those of other floating structures, which effectively suppresses low-frequency motions. However, this design also introduces complex, high-order dynamic responses under combined wind, wave, and current loads, leading to highly nonlinear tension fluctuations in the tendons that challenge structural reliability. Accurate, real-time prediction of extreme mooring tensions is therefore critical for ensuring safety and informing both design and operation. While conventional recurrent neural networks (RNNs)—such as standard RNNs, Long Short-Term Memory networks (LSTMs), and Gated Recurrent Units (GRUs)—are commonly used for time-series forecasting in offshore engineering, they often struggle to capture long-range dependencies within dynamic sequences, particularly during extreme loading events. To overcome this limitation, we propose a novel hybrid deep learning model that integrates a GRU with a multi-head attention (MHA) mechanism. This integration enables the model to dynamically assess the importance of past states over extended histories, thereby explicitly resolving long-range dependencies while preserving temporal detail. Using the MIT/NREL 5 MW TLP as a case study, we generated high-fidelity mooring tension data through coupled aero-hydro-mooring-servo simulations under various environmental conditions, formatting the data using a sliding-window approach. The proposed MHA-GRU model was trained and tested on this data, demonstrating significantly better predictive accuracy compared to a baseline GRU. These results indicate the model's high suitability for real-time mooring load forecasting and structural health monitoring of TLP FOWTs.*

**Keywords:** TLP, FOWT, Mooring, Machine-Learning, GRU, Multi-Head Attention

## 1. INTRODUCTION

The transition to renewable energy is essential for achieving global carbon neutrality, and offshore wind power stands as a pivotal contributor [1]. Deep-sea regions, with their abundant wind resources, offer particularly vast potential for scaling up clean energy. Harnessing this potential depends on floating offshore wind turbines (FOWTs), whose stability in chaotic seas relies on robust mooring systems. However, extreme environmental forces—such as intense winds, waves, and currents—can induce critical loads on these moorings, risking line failure [2]. To mitigate this, engineers typically incorporate large safety margins in design. Yet this conservative approach significantly increases costs, with mooring systems alone accounting for over 20% of a FOWT's total expense, challenging the project's overall economic viability.

Accurate real-time prediction of mooring line tension is vital for ensuring structural integrity, operational safety, and cost-effectiveness. Traditional numerical methods, such as coupled aero-hydro-mooring simulations [3,4], rely on iterative solutions involving potential flow theory or high-fidelity Navier-Stokes-based models. While effective for controlled scenarios, these methods face limitations in real-world applications due to indeterminate boundary conditions and high computational costs, preventing real-time forecasting [5]. Moreover, FOWT operators require tension predictions with a lead time of several seconds to enable proactive control strategies, such as dynamic adjustments to turbine operations, which can enhance safety and allow for reduced design margins.

In recent years, artificial neural networks (ANNs) have emerged as promising tools for real-time prediction tasks in ocean engineering, leveraging their ability to model nonlinear relationships in high-dimensional data. Recurrent neural networks (RNNs), including long short-term memory (LSTM) and gated recurrent unit (GRU) models [6], have been applied to forecast motions and mooring tensions in floating structures [7]. For instance, studies have demonstrated the effectiveness of bidirectional LSTM (Bi-LSTM) [8] and multi-input LSTM (MI-LSTM) [9] in capturing temporal dependencies using platform motion and environmental data. However, existing research has primarily focused on semi-submersible and spar-type platforms, with limited attention to tension-leg platforms (TLPs).

TLP-based FOWTs are gaining attention due to their advantages in hydrodynamic performance and smaller seabed footprint [10]. However, the taut mooring system of TLPs results in high global stiffness, making the platform more sensitive to mooring line integrity. A single line failure could compromise the entire system's safety, underscoring the need for reliable real-time tension monitoring. Conventional RNN architectures often struggle to capture long-range dependencies in time-series data, particularly under extreme loading events. To address this gap, we propose a hybrid model combining GRU with a multi-head attention (MHA) mechanism. The MHA enhancement allows the model to dynamically weigh the importance of historical states over extended sequences, effectively resolving long-range dependencies while maintaining temporal precision. The main contributions of this work are as follows:

- We introduce a GRU-MHA model tailored for real-time mooring tension prediction in TLP FOWTs, capable of handling complex environmental misalignments.
- The model is validated using high-fidelity simulation data from a coupled aero-hydro-elastic analysis of the MIT/NREL 5MW TLP benchmark, demonstrating superior performance over a baseline GRU, especially for longer input sequences.
- We provide a detailed analysis of the model's ability to capture critical dynamic features, such as low-frequency resonances and peak tension events, which are essential for structural health monitoring.

The remainder of this paper is organized as follows. Section 2 describes the methodology, including the GRU and MHA architectures, and the integrated GRU-MHA approach. Section 3 presents the case study setup, including the TLP FOWT model and data generation process. Section 4 discusses the results, highlighting the predictive accuracy and computational efficiency of the proposed model. Finally, Section 5 concludes the paper with insights and future research directions.

## 2. METHODOLOGY

### 2.1 Gated Recurrent Units (GRU)

The Gated Recurrent Unit (GRU) [11] is a specialized type of Recurrent Neural Network (RNN) architecture designed to model sequential or time-series data. For engineering tasks like forecasting the dynamic response (e.g., tension, position) of mooring lines in floating offshore wind turbines (FOWTs), GRU networks are exceptionally well-suited. They excel at learning complex temporal dependencies from multivariate input sequences—such as historical time series of wind speed, wave height, wave period, and turbine operational data—to predict future loads and motions. The core innovation of the GRU is its gating mechanism, which allows the model to adaptively control the flow of information. It decides what historical information to retain(memory) and what new information to incorporate, effectively solving the vanishing gradient problem common in standard RNNs and enabling the learning of long-term dependencies critical for simulating slow-drift oscillations and storm-level events in mooring systems.

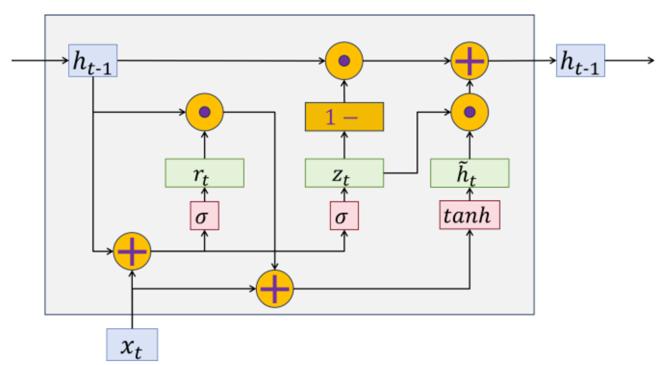



**FIGURE 1:** ILLUSTRATION OF A GRU CELL

The mathematical formulation of the GRU model can be expressed by the following equations:

$$z_t = \sigma\left(W_z \cdot \left[h_{\{t-1\}}, x_t\right] + b_z\right) \tag{1}$$

$$r_t = \sigma\left(W_r \cdot \left[h_{\{t-1\}}, x_t\right] + b_r\right) \tag{2}$$

$$\widetilde{h_t} = \tanh\left(W_h \cdot \left[r_t \odot h_{\{t-1\}}, x_t\right] + b_h\right) \tag{3}$$

$$h_t = (1 - z_t) \odot h_{\{t-1\}} + z_t \odot \widetilde{h_t} \tag{4}$$

In the Gated Recurrent Unit (GRU) formulation from Eq. (1) to Eq. (4), the variable $t$ denotes the current timestep, with the input vector $x_t$ representing concurrent measurements such as wind speed, wave height, and wave period. The hidden state from the previous timestep, $h_{t-1}$, encodes historical sequence information, while the update gate output $z_t$ (bounded in [0,1] by the sigmoid activation function σ) determines the degree to which $h_{t-1}$ is retained, and the reset gate output $r_t$ (also in [0,1]) controls its influence on the candidate hidden state $\widetilde{h_t}$. Trainable parameters include weight matrices $W_z$, $W_r$, and $W_h$, along with bias vectors $b_z$, $b_r$, and $b_h$. Vector concatenation, denoted by $[,]$, combines inputs for gate computations, and the product $\odot$ performs element-wise multiplication. The hyperbolic tangent function $tanh$ maps values to [-1,1], used in computing $\widetilde{h_t}$, which is a proposed new memory based on the current input and a gated version of the past. The final hidden state $h_t$ is derived as a combination of $h_{t-1}$ and $\widetilde{h_t}$, serving as the output for timestep $t$ and the memory for the next step.

## 2.2 Multi-Head Attention (MHA)

The Multi-Head Attention (MHA) mechanism [12] is a core architectural component of the Transformer model [12], designed to model relationships and dependencies within a sequence—regardless of the distance between elements. In the context of forecasting tasks for FOWT mooring lines, MHA excels at discovering complex, non-linear interactions between different parameters (e.g., wind, wave, current) and across different time steps. Its fundamental idea is to allow the model to jointly attend to information from different representation subspaces at different positions. Instead of performing a single attention function, MHA runs multiple, parallel "attention heads." This enables the model to simultaneously capture various types of relationships—for instance, one head might focus on short-term, high-frequency wave correlations, while other attends to long-term trends in wind speed or resonant platform motions. This parallel, multi-perspective analysis makes MHA exceptionally powerful for learning from the intricate, multi-physics dynamics that govern mooring system responses.

The mathematical formulation of the MHA mechanism can be expressed by the following equations. First, the scaled dot-product attention is defined as:

$$Attention(Q, K, V) = softmax\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right) \cdot V \tag{5}$$

Then the Linear Projections for Queries, Keys, and Values are:

$$Q = X \cdot W_Q \tag{6}$$

$$K = X \cdot W_K \tag{7}$$

$$V = X \cdot W_V \tag{8}$$

Thus, the complete Multi-Head Mechanism can be written as:

$$head_i = Attention\left(X \cdot W_{Q_i}, X \cdot W_{K_i}, X \cdot W_{V_i}\right) \tag{9}$$

$$\begin{aligned} &MultiHead(X) \\ &= \boldsymbol{Concat}(head_1, head_2, \ldots, head_h) \cdot W_O \end{aligned} \tag{10}$$

Within the Multi-Head Attention architecture, the variable definitions are articulated as follows: where $X$ denotes the input matrix containing sequential feature vectors; $Q$ (Query) represents the projection capturing information requests; $K$ (Key) signifies the identifiers for similarity computation; $V$ (Value) corresponds to the content-bearing elements; $W_Q$, $W_K$, and $W_V$ indicate the learnable projection matrices generating $Q$, $K$, and $V$ representations respectively; $d_k$ defines the dimensionality of Query and Key vectors; h specifies the number of parallel attention heads; $W_{Q_i}$, $W_{K_i}$, $W_{V_i}$ refer to head-specific projection matrices for the i-th attention head; $head_i$ indicates the output of the $i$-th attention mechanism; $\boldsymbol{Concat}$ denotes the operation combining all head outputs; and $W_O$ represents the final projection matrix synthesizing the concatenated outputs into a unified representation.

## 2.3 GRU-MHA Network

In the present work, a hybrid architecture is developed that effectively leverages the complementary strengths of the Gated Recurrent Unit (GRU) and the Multi-Head Attention (MHA) mechanism. Within this framework, the GRU module is responsible for encoding the temporal sequences, capturing both short-term dynamics and evolving patterns in the input data. Subsequently, the MHA module operates on the GRU's encoded sequence representations, enabling the model to explicitly capture long-range contextual dependencies by adaptively attending to relevant time steps across extended historical horizons. This synergistic integration allows the model to simultaneously preserve fine-grained temporal details through the GRU while effectively resolving long-range interactions via the MHA, resulting in enhanced predictive performance for mooring tension forecasting. An illustration of the GRU-MHA model is provided in Fig. 2.

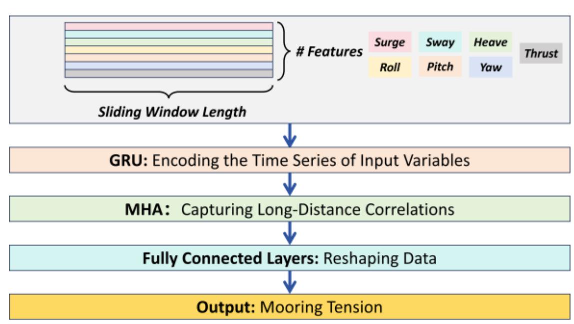



**FIGURE 2:** ILLUSTRATION OF THE GRU-MHA MODEL

Specifically, the GRU processes the sequential input data, such as time-series sensor measurements, to capture complex temporal dynamics and dependencies over time. Its gating mechanism allows it to effectively remember important long-term trends and forget irrelevant information, making it well-suited for modeling the evolving environmental and response history of a floating system. Following this, the outputs are passed to a multi-head attention layer with 4 parallel attention heads. This layer enables the model to simultaneously analyze the processed sequence from four different representational perspectives or subspaces. For instance, one head might focus on correlating peak tensions with specific wave conditions, while another might discover relationships between slow-drift motions and wind gust patterns. By attending to different aspects of the GRU's encoded sequence, the attention layer identifies and emphasizes the most critical interactions and time steps for the final forecast. Finally, the insights from all four heads are synthesized to produce a robust prediction.

## 3. TEST CASE DESCRIPTION

### 3.1 The MIT/NREL 5MW TLP FOWT

The Tension-Leg Platform (TLP) floating offshore wind turbine based on the MIT/NREL 5MW [13] benchmark represents a standard reference numerical model widely used in deep-water wind energy research. This system integrates the well-established NREL 5MW [14] reference wind turbine—a three-bladed, upwind horizontal-axis machine with a rated power of 5 MW and a rotor diameter of 126 m—with a TLP floating substructure. The TLP design employs a vertically taut mooring system, which provides high global stiffness and excellent station-keeping performance, effectively suppressing low-frequency motions while introducing complex dynamic responses under combined environmental loads. The platform features a cylindrical hull with a draft of 47.89 m and a diameter of 18 m, ballasted to achieve static stability, and is stationed at a water depth of 200 m. Its mooring system consists of 8 tension legs arranged in a symmetric pattern, ensuring restoring moments in pitch, roll, and heave, as shown in Fig. 3.

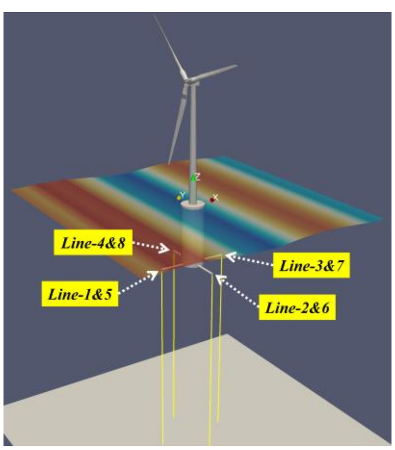



**FIGURE 3:** SETUP OF THE MIT/NREL 5MW TLP FOWT

### 3.2 Data Generation

A series of simulations were conducted with misaligned environmental conditions. The wave direction was fixed at 0 degrees. The wind direction (*wdd*) and current direction (*crd*) were varied from 0 to 180 degrees in 30-degree increments, creating a comprehensive matrix of misalignment cases. The simulated environmental conditions comprised three components: a turbulent wind field with a mean hub-height (90 m) velocity of 12 m/s; an irregular wave field characterized by a JONSWAP spectrum (significant wave height = 6.1 m, peak period = 11.3 s, shape parameter = 1.8); and a uniform current with a speed of 0.52 m/s. An illustration of the numerical configuration is given in Fig. 4.

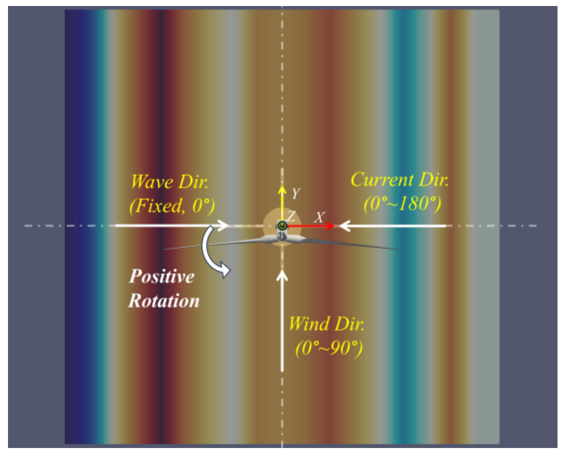



**FIGURE 4:** DEFINITION OF WAVE-WIND-CURRENT DIRECTIONS

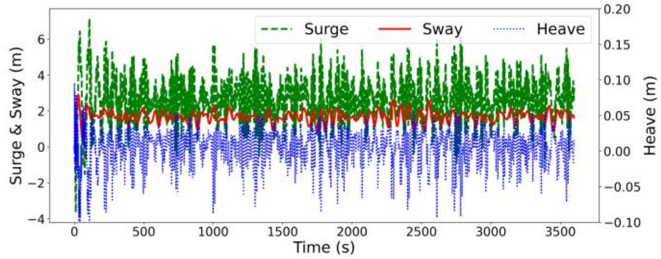
In the current study, the numerical simulations assume aligned The JONSWAP random wave spectrum is defined as follows [15]:

$$S(f) = \alpha \frac{H_s^2}{T_p^4 f^5} exp\left[-\frac{5}{4}(T_p f)^{-4}\right] \gamma^{exp\left[-\frac{(T_p f-1)}{2\sigma^2}\right]} \quad (11)$$

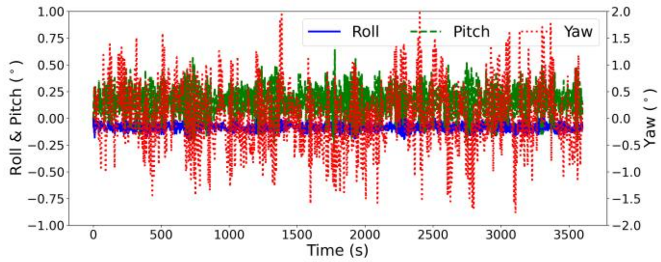
where $\alpha$ is the Phillips constant, $T_p$ is the peak wave period, $H_s$ is the significant wave height, $f$ is the wave frequency, and $\gamma$ is the shape parameter, which defines the sharpness of the spectral profile. The value of $\sigma$ is determined as follows:

$$\sigma = \begin{cases} 0.07, f \le f_p \\ 0.09, f > f_p \end{cases} \quad (12)$$

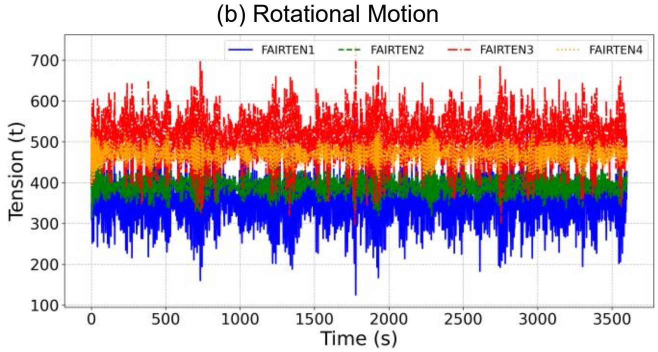
The simulation results are provided in Fig. 5.



(a) Translational Motion



(b) Rotational Motion



(c) Mooring Tension

**FIGURE 5:** SIMULATION RESULTS FOR THE 0°-30°-60° WAVE-WIND-CURRENT DIRECTIONS

## 4. RESULTS AND DISCUSSION

### 4.1 Definition of Neural Network Mapping

The objective of the neural network mapping in this work is to predict fairlead mooring tensions. This is achieved by establishing a functional relationship between the system's historical states and a future tension value. Specifically, the networks take historical sequences of the platform's six-degree-of-freedom motion and rotor thrust as inputs. The corresponding mapping is defined as:

$$Mooring\ Tensions^{t} = \mathcal{NN}(motion^{his}, thrust^{his}) \qquad (13)$$

In this formulation, $\mathcal{NN}$ represents the nonlinear mapping learned by the network, the superscript $t$ denotes the time instant for which tension is predicted, and *his* signifies the input history used for the prediction.

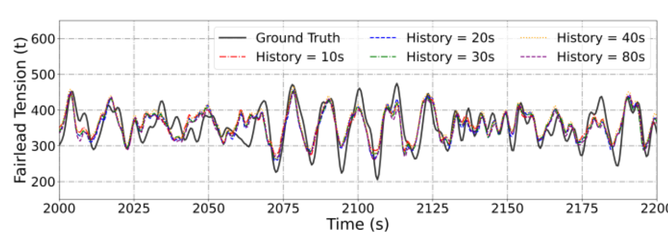
### 4.2 Model Training

The proposed GRU-MHA model was trained over 1000 epochs using the Adam optimizer with a learning rate of $1 \times 10^{-4}$ and a batch size of 1000, balancing predictive accuracy and computational efficiency for real-time mooring tension forecasting. The architecture comprises a 7-layer GRU network to capture temporal dependencies, a 4-head multi-head attention mechanism for multi-perspective sequence analysis, and 3 fully connected hidden layers with ReLU activation, culminating in a linear output layer for continuous tension prediction. Training data were derived from high-fidelity coupled simulations of the MIT/NREL 5MW TLP under misaligned environmental conditions, preprocessed into sequential samples via sliding windows and partitioned into training, validation, and test sets. Model convergence was monitored using mean absolute error (MAE) as the loss metric, with early stopping implemented after 50 epochs of validation loss plateau to ensure robustness and prevent overfitting.

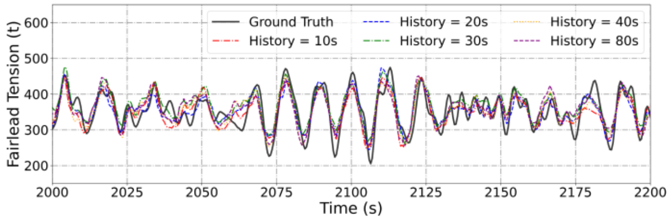
### 4.3 Model Evaluation

First, it should be mentioned here that for clarity, subsequent illustrations focus solely on the one of the eight mooring lines experiencing the highest load. The predictive performance of the GRU and GRU-MHA models is comprehensively evaluated using both training and testing datasets, as visually summarized in Figs. 6 through 8. Fig. 7 illustrates the models' performance on a representative training set, while Figs. 6 and 8 depict their generalization capabilities on two distinct testing sets under varying environmental misalignment scenarios. A critical hyperparameter explored in this analysis is the input sequence length, or history length (denoted as History or His. in the subsequent figures), which directly influences the models' ability to capture temporal dependencies (note that in the current study, the lead time of prediction is fixed as 10 seconds). Consequently, each figure presents results from models trained with different history lengths, ranging from 10 to 80 seconds, to systematically assess the impact of extended historical context on prediction accuracy.

The time-series plots demonstrate that both architectures achieve commendable overall performance, closely tracking the general trends of the actual mooring tension data across most operating conditions. The predicted curves exhibit strong alignment with the ground truth in terms of phase and amplitude for low-frequency components. However, a detailed inspection reveals notable discrepancies in the high-frequency regime, where both models struggle to accurately capture rapid tension fluctuations. The GRU-MHA model consistently shows superior performance in attenuating these high-frequency errors compared to the standalone GRU, owing to its attention mechanism's ability to focus on critical temporal features.

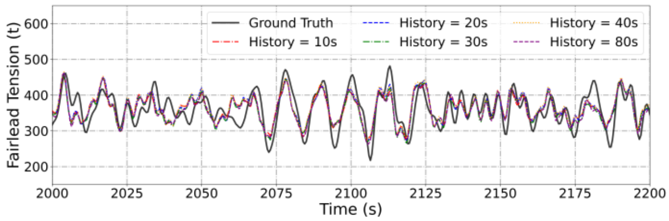


(a) GRU Predictions

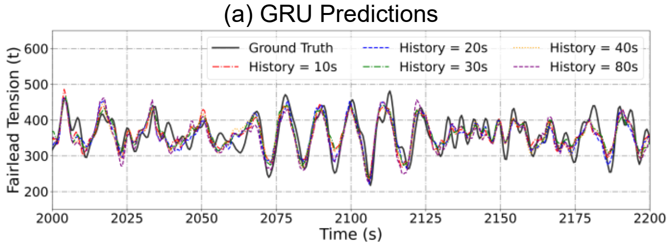



(b) GRU-MHA Predictions

**FIGURE 6:** NN PREDICTIONS FOR THE 0°-0°-180° WAVE-WIND-CURRENT DIRECTIONS (TESTING SET) UNDER DIFFERENT HISTORY LENGTHS

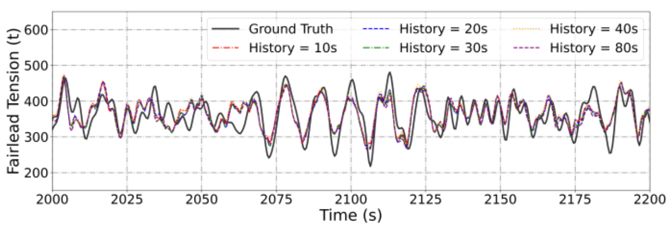



(a) GRU Predictions

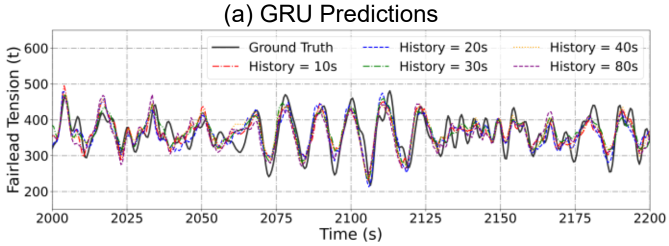



(b) GRU-MHA Predictions

**FIGURE 7:** NN PREDICTIONS FOR THE 0°-30°-120° WAVE-WIND-CURRENT DIRECTIONS (TRAINING SET) UNDER DIFFERENT HISTORY LENGTHS




(a) GRU Predictions




(b) GRU-MHA Predictions

**FIGURE 8:** NN PREDICTIONS FOR THE 0°-60°-180° WAVE-WIND-CURRENT DIRECTIONS (TESTING SET) UNDER DIFFERENT HISTORY LENGTHS

The evaluation further underscores the importance of history length optimization. Excessively short histories (e.g., 10 seconds) lead to underfitting of long-period dynamics, while overly long sequences (e.g., 80 seconds) introduce noise and computational inefficiency without substantive accuracy gains. The GRU-MHA hybrid demonstrates enhanced robustness to history length variations, effectively leveraging extended sequences to improve predictions, whereas the GRU model exhibits performance saturation and gradual degradation beyond a 20-second history. Overall, the figures corroborate that the proposed GRU-MHA architecture offers a balanced compromise between low-frequency trend accuracy and high-frequency detail preservation, making it particularly suitable for real-time mooring tension forecasting in dynamic offshore environments.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - y_i^{pred}\right| \tag{13}$$

$$RAE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - y_i^{pred}}{y_i}\right| \tag{14}$$

This trend is quantitatively supported by the MAE and RAE (defined by Eq. (13) and Eq. (14), respectively) metrics reported in Table 1. Overall, the GRU-MHA hybrid architecture demonstrates consistently superior accuracy compared to the standalone GRU model, with lower mean absolute error (MAE) and relative absolute error (RAE) values observed throughout all tested sequence lengths. This performance advantage is particularly evident in the 20-40 second history range, where the GRU-MHA model achieves its optimal precision with MAE values around 25.3 t ($RAE \approx 7.2\%$), representing approximately 12-15% improvement over the baseline GRU model. The enhanced performance can be attributed to the multi-head attention mechanism's ability to dynamically weight important temporal patterns across extended sequences, enabling more effective learning of long-range dependencies in the mooring tension dynamics.

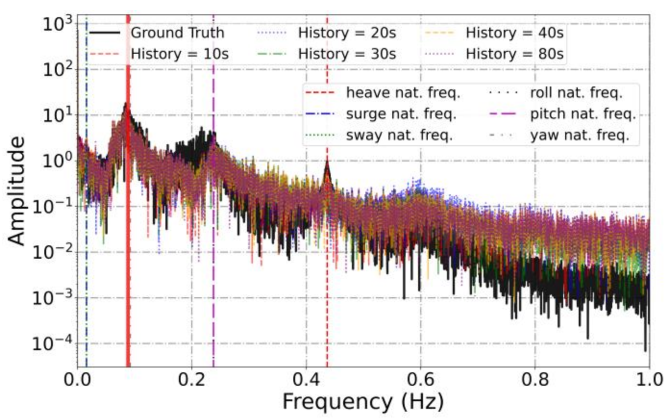
A detailed examination of the history length optimization reveals fundamentally different behaviors between the two architectures. The standard GRU model shows limited capacity to leverage extended historical contexts, achieving its best performance at 20 seconds ($MAE = 28.4$ t, $RAE = 8.0\%$) with progressive degradation as history length increases beyond this point. This pattern indicates the model's inherent limitation in extracting useful information from longer sequences, ultimately saturating its learning capability and potentially introducing noise that compromises predictive accuracy. In contrast, the GRU-MHA model exhibits robust performance improvements as history length increases from 10 to 40 seconds, maintaining stable optimal performance between 30-40 seconds before experiencing degradation at 80 seconds. This behavior demonstrates the attention mechanism's effectiveness in filtering redundant information while amplifying critical long-range temporal relationships.

**TABLE 1:** MAE / RAE ACROSS THE TESTING DATA SET.

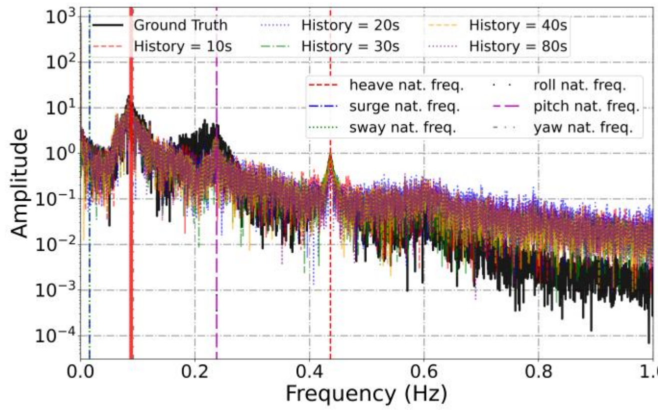
| His Lenth | GRU | GRU-MHA |
| --- | --- | --- |

| | | |
|---|---|---|
| His = 10s | 29.7 t / 8.4% | 26.8 t / 7.6% |
| His = 20s | 28.4 t / 8.0% | 25.7 t / 7.3% |
| His = 30s | 28.7 t / 8.1% | 25.3 t / 7.2% |
| His = 40s | 28.9 t / 8.2% | 25.3 t / 7.2% |
| His = 80s | 29.4 t / 8.3% | 28.3 t / 8.0% |

The comparative analysis underscores a fundamental distinction in how each architecture processes temporal information. The GRU's performance plateau and subsequent degradation beyond 20 seconds history suggests diminishing returns from longer input sequences, potentially increasing computational burden without corresponding accuracy gains. This limitation stems from the GRU's fixed-length memory mechanism, which struggles to maintain relevant information across extended time horizons. The incorporation of multi-head attention effectively addresses this constraint by enabling the model to selectively focus on strategically important time steps regardless of their temporal distance, thereby unlocking the predictive value embedded in longer historical contexts. The performance degradation observed in both models at the 80-second history length suggests the existence of an optimal context window beyond which additional historical data introduces more noise than signal, though the GRU-MHA's superior degradation resistance highlights its enhanced robustness to suboptimal hyperparameter selection. These findings collectively validate the MHA enhancement as a critical innovation for mooring tension forecasting, particularly for TLP systems where load dynamics involve complex interactions across multiple time scales.



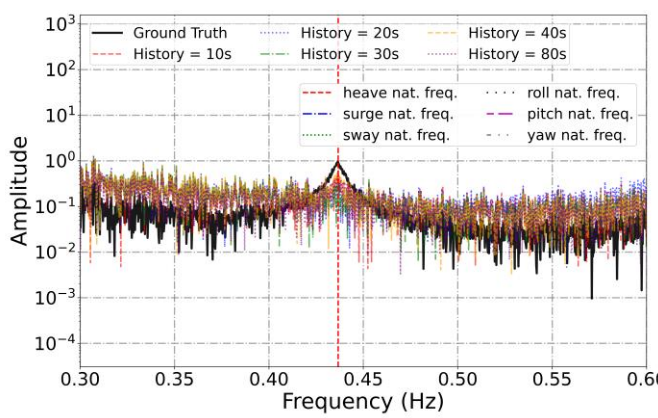
(a) GRU Predictions



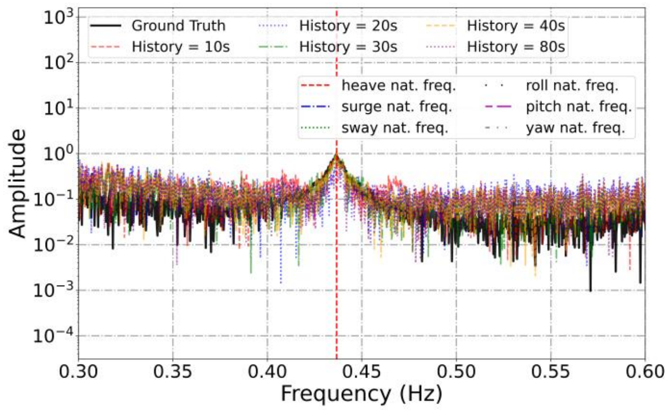
(b) GRU-MHA Predictions

**FIGURE 9:** FREQUENCY-DOMAIN CHARACTERISTICS OF THE NN-PREDICTED RESPONSES UNDER WAVE-WIND-CURRENT DIRECTIONS OF 0°-60°-180° (TESTING SET) FOR DIFFERENT HISTORY LENGTHS

The frequency-domain analysis of the prediction errors, as illustrated in Fig. 9, reveals that both the GRU and GRU-MHA models achieve high accuracy in the low-frequency range, successfully capturing key structural natural frequencies such as the wave peak period (Tp), pitch natural frequency, and heave natural frequency. However, both models exhibit amplified errors in the high-frequency region above 0.5 Hz, likely due to numerical oscillations introduced by the neural network architectures during the training process. This high-frequency deviation suggests a common limitation in data-driven approaches for modeling rapid transient dynamics.



(a) GRU Predictions




(b) GRU-MHA Predictions

**FIGURE 10:** FREQUENCY-DOMAIN DETAILS OF THE NN-PREDICTED RESPONSES FOR THE WAVE-WIND-CURRENT DIRECTIONS OF 0°-60°-180° (TESTING SET) UNDER DIFFERENT HISTORY LENGTHS.

A detailed examination of the heave natural frequency region in Fig. 10 demonstrates the superior capability of the GRU-MHA model. While the standard GRU fails to resolve a distinct spectral peak at this critical frequency, the GRU-MHA variant successfully captures the sharp resonance characteristic, indicating its enhanced ability to model structural dynamics through the multi-head attention mechanism. This improved frequency-domain performance underscores the value of incorporating attention mechanisms for capturing complex structural responses in mooring tension predictions.

## 5. CONCLUSION

This study developed and validated a novel hybrid deep learning framework that integrates Gated Recurrent Units with Multi-Head Attention (GRU-MHA) for real-time mooring tension prediction in Tension-Leg Platform (TLP) floating offshore wind turbines (FOWTs). The research establishes a comprehensive methodology for processing complex time-series data from coupled wind-wave-current environments, employing advanced neural network architectures to address the critical challenge of predicting extreme mooring loads in offshore renewable energy systems.

The investigation reveals several significant findings regarding mooring tension prediction performance. The GRU-MHA model consistently outperforms the baseline GRU architecture across all evaluated history lengths with a fixed lead time of prediction of 10 seconds, achieving approximately 12-15% improvement in mean absolute error for optimal sequence lengths of 30-40 seconds. This performance advantage is particularly pronounced in capturing peak tension events and low-frequency resonant responses, which are critical for structural reliability assessments. The multi-head attention mechanism proves essential for modeling long-range dependencies in the dynamic response data, enabling the network to effectively correlate distant temporal events while maintaining precision in short-term predictions. Furthermore, the model demonstrates robust performance across various environmental misalignment conditions, showing particular strength in resolving complex hydrodynamic interactions that characterize TLP behavior under combined loading scenarios.

Despite these promising results, several limitations warrant consideration and present directions for future research. Future work will focus on several key areas: experimental validation using field measurement data from operational TLP installations, extension of the methodology to multi-line tension prediction and coupled platform response forecasting, investigation of transfer learning approaches for adapting pre-trained models to new site conditions, and development of more efficient attention mechanisms to reduce computational overhead for real-time applications. Additionally, exploring the integration of physics-based constraints into the learning process could enhance model interpretability and generalization capability across varying environmental conditions.

## ACKNOWLEDGEMENTS

This work was supported by the National Key Research and Development Program of China (2025YFF0515700), to which the authors are most grateful.

## REFERENCES

[1] 2024, *World Energy Outlook 2024*, International Energy Agency, Paris, France.

[2] Jiang, Z., 2025, "Mooring Design for Floating Wind Turbines: A Review," Renew. Sustain. Energy Rev., **212**, p. 115231.

[3] Wang, Y., Chen, H., Koop, A., and Vaz, G., 2022, "Hydrodynamic Response of a FOWT Semi-Submersible under Regular Waves Using CFD : Verification and Validation," Ocean Eng., **258**(January), p. 111742.

[4] Xu, S., Zhuang, T., Zhao, W., and Wan, D., 2023, "Numerical Investigation of Aerodynamic Responses and Wake Characteristics of a Floating Offshore Wind Turbine under Atmospheric Boundary Layer Inflows," Ocean Eng., **279**(April), p. 114527.

[5] Ye, M., Cao, K., Wang, J., and Wan, D., 2026, "Mooring Tension Prediction of Floating Offshore Wind Turbines Using End-to-End Convolutional Neural Network Models with Attention Mechanisms," Ocean Eng., **346**, p. 123939.

[6] Xie, Y., Tang, H., and Low, Y. M., 2023, "Deep Gated Recurrent Unit Networks for Time-Domain Long-Term Fatigue Analysis of Mooring Lines Considering Wave Directionality," Ocean Eng., **284**, p. 115244.

[7] Deng, S., Ning, D., and Mayon, R., 2024, "The Motion Forecasting Study of Floating Offshore Wind Turbine Using Self-Attention Long Short-Term Memory Method," Ocean Eng., **310**, p. 118709.

[8] Yin, J., Ding, J., Yang, Y., Yu, J., Ma, L., Xie, W., Nie, D., Bashir, M., Liu, Q., Li, C., and Li, S., 2025, "Wave-Induced Motion Prediction of a Deepwater Floating Offshore Wind Turbine Platform Based on Bi-LSTM," Ocean Eng., **315**, p. 119836.

[9] Tang, X., Huang, W., Li, X., and Ma, G., 2025, "Prediction of Mooring Tension of Floating Offshore Wind Turbines by CNN-LSTM-ATT and Chebyshev Polynomials," Ocean Eng., **331**, p. 121327.

[10] Wang, Y., Yao, T., Zhao, Y., and Jiang, Z., 2025, "Review of Tension Leg Platform Floating Wind Turbines: Concepts, Design Methods, and Future Development Trends," Ocean Eng., **324**, p. 120587.

[11] Chung, J., Gulcehre, C., Cho, K., and Bengio, Y., 2014, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling."

[12] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I., 2017, "Attention Is All You Need," *31st Conference on Neural Information Processing Systems (NIPS 2017)*.

[13] Matha, D., 2010, *Model Development and Loads Analysis of an Offshore Wind Turbine on a Tension Leg Platform with a Comparison to Other Floating Turbine Concepts*, NREL/SR-500-45891, 973961, National Renewable Energy Laboratory, Colorado, USA.

[14] Jonkman, J., Butterfield, S., Musial, W., and Scott, G., 2009, *Definition of a 5-MW Reference Wind Turbine for Offshore System Development*.

[15] Mazzaretto, O. M., Menéndez, M., and Lobeto, H., 2022, "A Global Evaluation of the JONSWAP Spectra Suitability on Coastal Areas," Ocean Eng., **266**, p. 112756.