SIN-CHN
SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2025, 22(2), 1235.
https://doi.org/10.62617/mcb1235

Article

Biomechanical data-driven prediction and analysis based on transformer model

Zheyang Yan[1], Wenchao Fan[2,*]

[1] Department of Physics and Information Engineering, Cangzhou Normal University, Cangzhou 061001, China
[2] Department of Computer Science and Engineering, Cangzhou Normal University, Cangzhou 061001, China
*** Corresponding author:** Wenchao Fan, czsygzc@126.com

CITATION

Yan Z, Fan W. Biomechanical data-driven prediction and analysis based on transformer model. Molecular & Cellular Biomechanics. 2025; 22(2): 1235.
https://doi.org/10.62617/mcb1235

ARTICLE INFO

Received: 20 November 2024
Accepted: 16 January 2025
Available online: 7 February 2025

Abstract: With the development of high-precision sensors and data acquisition equipment, biomechanical data presents high-dimensional, strong time-series dependence and nonlinear characteristics, and it is difficult for traditional physical modeling and statistical methods to process such data efficiently and accurately. The purpose of this paper is to build a biomechanical data-driven prediction framework based on Transformer model, and realize high-precision prediction by deeply mining the time series characteristics of data, which provides theoretical support and practical application value for medical diagnosis, rehabilitation monitoring and sports science. In terms of methods, this paper preprocesses biomechanical data such as joint angle, electromyography (EMG) and joint stress, and designs a time series prediction framework based on the self-attention mechanism of Transformer model. Through the simulation experiment, five indexes, namely mean square error (MSE), mean absolute error (MAE), determination coefficient (R^2), prediction time and Pearson correlation coefficient, are selected to evaluate and compare the performance of the model. The experimental results show that the Transformer model is superior to the traditional LSTM, GRU and ARIMA models in all kinds of biomechanical data prediction tasks: MSE is 0.0152, R^2 is as high as 0.982, and the prediction time is only 0.76 s. In addition, Pearson correlation coefficient is close to 1 in different data types, which verifies the high consistency between the predicted results of the model and the real values. The conclusion of this paper shows that the Transformer model can effectively capture the global spatio-temporal characteristics of biomechanical data, and has high precision, high efficiency and strong generalization ability, which provides new technical means and theoretical support for biomechanical data-driven analysis and application.

Keywords: biomechanical data; transformer model; data driven; high precision prediction

1. Introduction

Biomechanics, as an important subject to study human motion and mechanical behavior, is widely used in medicine, sports science, rehabilitation engineering and other fields. With the development of sensor technology and data acquisition equipment (such as high-precision mechanical sensor, motion capture system, electromyography measurement equipment, etc.), the accuracy and scale of biomechanical data acquisition have been significantly improved. However, this kind of data usually has the characteristics of high dimension, strong time series dependence and nonlinearity. Traditional physical models and statistical methods have many challenges in dealing with this kind of data, such as high modeling complexity, insufficient generalization ability and low computational efficiency [1].

In recent years, with the help of machine learning and deep learning technology, data-driven methods have shown remarkable advantages in complex data modeling and prediction [2]. Especially in the task of time series data prediction, methods such as recurrent neural network (RNN) and long-term and short-term memory network (LSTM) have achieved good results [3]. However, due to its recursive structure, this kind of model has the problems of difficult parallel calculation and limited long-term dependence modeling ability, which is difficult to meet the accurate prediction requirements of high-complexity biomechanical data.

Aiming at the above problems, the Transformer model provides a new solution for the time series prediction of biomechanical data. Transformer model overcomes the limitations of traditional time series model by using Self-Attention mechanism, and can capture long-term dependencies and realize efficient parallel computing. In this paper, biomechanical data is taken as the research object, and a prediction framework based on Transformer model is constructed, aiming at mining the deep time series characteristics of data, realizing high-precision biomechanical data-driven prediction and analysis, and providing theoretical support and practical application value for medical diagnosis, rehabilitation monitoring and sports science.

2. Theoretical background

2.1. Biomechanical data-driven prediction

Biomechanics is a subject that studies the mechanical characteristics and motion laws of biological systems (such as human body and animals), covering the interactive mechanical relations among complex systems such as bones, muscles and joints. Traditional biomechanical analysis usually relies on physical modeling, such as finite element analysis (FEA) or solving dynamic equations [4]. These methods are excellent in analyzing linear and simple structures, but in complex biological systems, it is difficult for traditional modeling methods to accurately describe the system behavior due to organizational heterogeneity, nonlinear material properties and multi-scale coupling characteristics [5,6].

Data-driven method learns hidden features and mapping relationships in biomechanical data through machine learning algorithm, skips physical modeling and directly models and predicts the system. This method is suitable for processing large-scale biomechanical data with strong individual differences. For example, based on the prediction of human gait data, the stress and motion trajectory of skeletal joints can be accurately estimated, which provides theoretical support for rehabilitation training and sports injury protection. In recent years, biomechanical analysis tools widely used in sports science and clinical settings, such as motion capture systems and wearable devices, have produced a large number of time series data, providing rich materials for data-driven prediction.

In the training of athletes, the prediction model based on biomechanical data can help coaches and athletes to evaluate the joint stress and muscle fatigue in real time, so as to optimize the exercise plan and reduce the risk of sports injury. In addition, in rehabilitation medicine, by analyzing the gait and muscle activity patterns of patients, the data-driven prediction model can provide strong support for

the formulation of personalized treatment programs and improve the rehabilitation effect.

Biomechanical data, such as EMG, joint angle and mechanical response, usually show the characteristics of time continuity and dynamic change. Therefore, the time series prediction model has become the core tool to solve biomechanical problems. Traditional time series models, such as autoregressive moving average model (ARIMA) and long-term memory network (LSTM), can capture the time correlation of data, but when dealing with long-term dependence and high-dimensional complex data, there are some problems such as low computational efficiency and insufficient generalization ability of models. The introduction of Transformer model provides a new way to solve this problem.

2.2. Introduction of transformer model

Transformer model was proposed by Vaswani et al. in 2017, and was originally applied to natural language processing (NLP) tasks. Different from the traditional recurrent neural network (RNN) and convolutional neural network (CNN), Transformer is completely based on the self-attention mechanism, which realizes the efficient modeling of the dependencies between elements in sequence data.

The core structure of Transformer model includes self-attention mechanism, feedforward neural network, residual connection and layer normalization. Specifically, the self-attention mechanism determines the importance of each element in the prediction process by calculating the attention weight between each element and other elements in the input sequence. This modeling method of global dependency makes up for the deficiency of traditional RNN that it is difficult to capture long-distance dependency when the time step increases [7].

Transformer's Multi-Head Attention mechanism further improves the modeling ability of the model for different subspace features. By calculating multiple attention heads in parallel, the model can pay attention to different dimensions and different position characteristics of input data, so that it can handle complex high-dimensional biomechanical data. Transformer introduces position information through Positional Encoding, so that the model can perceive the time sequence of input data, which is an indispensable function in time series prediction [8].

Compared with traditional time series networks such as LSTM, the advantages of Transformer are mainly reflected in the following points: high computational efficiency, because of removing the circular structure, Transformer can realize parallel computing and greatly improve the training and reasoning speed. Long-term dependency modeling ability is strong. Through self-attention mechanism, Transformer can capture the long-distance dependency between data globally. The model has good expansibility, and Transformer has flexible structural design, which can adjust the model parameters, such as the number of layers and heads, according to the task requirements [9,10].

In the field of biomechanical data prediction, the introduction of Transformer model provides a new idea for complex time series data modeling. Through the self-attention mechanism, the model can effectively capture the potential spatio-temporal characteristics in biomechanical data, and accurately predict the motion state, joint

stress and muscle activation signals. For example, in the field of robot control and motion analysis, Transformer is used to analyze the stress of robot joints, optimize the motion control strategy of robots, and even realize autonomous motion in complex environments. In addition, the potential applications of Transformer in clinical rehabilitation and sports medicine, such as gait analysis and muscle activation pattern recognition, are constantly being expanded and verified. With the development of technology, the application field of Transformer model is expanding to more biomechanical research fields, which provides more accurate prediction and decision support.

3. Transformer model

3.1. Data preprocessing

Data preprocessing is a key step to ensure the performance of biomechanical data-driven prediction model [11,12]. Biomechanical data are widely available, including human motion capture data, electromyography (EMG), joint force and mechanical response data. These data usually come from laboratory equipment (such as motion capture systems, surface electromyography collectors, pressure sensors, etc.) and wearable devices or real-time sensors (such as smart sports watches and inertial measurement units (IMU), etc.). These data usually have the characteristics of high dimension, nonlinearity, time series and noise interference, and there are some differences between different sensors and data sources. In order to make the Transformer model deal with these complex data effectively and ensure the data quality and consistency, the following preprocessing work is usually needed:

1) Data Normalization:

Because the dimension and range of biomechanical data are quite different, the numerical range of each feature may be different by several orders of magnitude [13]. The joint angle data may fluctuate in the range of −180 to 180 degrees, while the amplitude of EMG signal is usually small, which may be concentrated between 0 and 1, and the joint force may be between tens and hundreds of Newtons. If the non-normalized data is directly input, the weight adjustment may be biased towards the characteristics with large values in the process of model training, which will affect the convergence efficiency and prediction accuracy of the model. Therefore, the normalization method is used to map the data to the same range, such as [0, 1] or a range with a mean value of 0 and a standard deviation of 1. In order to further ensure the data quality, the following normalization formula is usually used:

$$x' = \frac{x - \mu}{\sigma} \tag{1}$$

where x represents the raw data, μ is the mean of the data, σ is the standard deviation, and x' is the normalized data. Different data sources (such as EMG signals and motion capture data) need to be normalized separately to ensure the consistency of model input data.

2) Time Series Segmentation:

Biomechanical data are usually recorded in the form of time series, reflecting the dynamic characteristics of human motion state or mechanical signals [14,15]. In

practical application, in order to adapt to the Transformer model, it is necessary to segment these continuous time series data. Usually, a fixed-length time window is used as the input of the model, and multiple groups of inputs and corresponding prediction targets are generated by sliding the window. Let the time window be T, the step size be s, then the input for the k time window is:

$$X_k = \{x_k, x_{k+1}, \cdots, x_{k+T-1}\}, y_k = x_{k+T} \tag{2}$$

where X_k represents the input sequence, and y_k is the target prediction value.

3) Outlier Processing:

In the process of data acquisition, biomechanical data may be affected by factors such as sensor accuracy, environmental interference or operational errors, resulting in abnormal values (such as prominent noise, jumping signals or unreasonable measured values). These abnormal values may interfere with the learning of the normal model, and even lead to the distortion of the training results. Therefore, it is necessary to detect and process the abnormal values in the data. Commonly used methods for handling outliers include:

Median filtering: replacing outliers with the median in the data window can smooth the signal and reduce the influence of outliers.

Rule of Three Sigma: Assuming that the data obeys normal distribution, the data points beyond the three standard deviation ranges $[\mu - 3\sigma, \mu + 3\sigma]$ of the mean μ are regarded as abnormal values, and they are deleted or corrected.

4) Missing value processing:

In the collection of biomechanical data, data is often missing due to equipment failure, signal loss or measurement interruption. If these missing values are ignored directly, the input data of the model may be incomplete, which will affect the prediction ability of the model. Therefore, the filling of missing values has become an important part of data preprocessing. Common treatment methods include:

Linear interpolation: By linear interpolation method, the missing values are filled with the linear estimated values of adjacent data points, so as to maintain the continuity of data.

Mean filling: the missing value is replaced by the mean value of this feature, which is suitable for scenes with no obvious time correlation.

According to the time series characteristics of biomechanical data, linear interpolation is usually considered as an effective filling method, because it can better maintain the time continuity and dynamic characteristics of the data.

5) Data set partition:

In order to reasonably evaluate the generalization performance and prediction ability of the model, it is necessary to partition the data set. Generally speaking, according to the proportion of 70%, 15% and 15%, the data is divided into training set, verification set and test set:

Training set: used for model training and parameter optimization, accounting for most of the data set.

Verification set: used to adjust the superparameter of the model and monitor over-fitting during training.

Test set: used to finally evaluate the performance of the model and test its prediction ability for unknown data.

When dividing data, attention should be paid to maintaining the consistency of feature distribution of all kinds of data, so as to avoid the distortion of evaluation results caused by distribution differences. All kinds of motion or signal patterns in the data set (such as joint angles at different motion stages and EMG signals of different muscle activities) should be distributed consistently in the training set, verification set and test set [16–18].

In order to ensure the quality and consistency of data, researchers usually use different data sources, such as published biomechanics data sets (such as Biomechanics Lab data sets or human motion analysis data sets), and combine the data collected in the laboratory. Data quality monitoring and preprocessing steps of multi-source data fusion are helpful to improve the robustness and generalization ability of the model.

3.2. Transformer structural design

Based on the characteristics of biomechanical data, such as time series, complexity and high dimension, a Transformer model suitable for time series prediction task is designed in this study. The whole structure of the model consists of input layer, Encoder and output layer, and the core mechanisms are self-attention mechanism and position coding. Through this structural design, the model can effectively capture the temporal and spatial dependence characteristics in time series data, and significantly improve the prediction accuracy and generalization ability [19,20]. The following will elaborate on the design of each part.

3.2.1. Input layer design

The main function of the input layer is to transform the original biomechanical time series data into a feature representation that the model can handle, and at the same time explicitly introduce time series information. Biomechanical data include multi-dimensional information such as joint angle, EMG signal and mechanical response. The input layer maps these high-dimensional time series data into a unified feature space through linear transformation, which provides a basis for subsequent processing.

Because the Transformer model itself does not have the natural ability to deal with sequence order, the input layer explicitly introduces time information through Positional Encoding. This position coding replaces the implicit time step processing method in traditional recurrent neural network (RNN) in Transformer, which enables the model to understand the time sequence of input data.

3.2.2. Design of encoder

Encoder is the core part of Transformer model, and its main task is to extract the depth features of input data and model the temporal dependency. The overall structure of the encoder is composed of multiple Encoder Layers stacked, and each encoder layer contains the following key components:

Self-attention mechanism is the core module of Transformer model, which is used to calculate the dependence between different time steps in input data. Different from the traditional recurrent neural network (RNN), the self-attention mechanism can pay attention to the features in the data globally by calculating the similarity between time steps, which makes the model perform particularly well in dealing with

long-term dependence problems. The function of self-attention mechanism is to allow the model to pay attention to the important features in time series at the same time, rather than just relying on the adjacent time step information. In the biomechanical data of joint stress, the model can simultaneously capture the stress change relationship between different time points through the self-attention mechanism, which is of great significance in long-term time series prediction.

On the basis of self-attention mechanism, Transformer introduces multi-head attention mechanism, that is, the same input data is calculated in parallel for many times, and different subspace features of the concerned data are calculated each time. Multi-attention mechanism can capture data features from different angles, which significantly enhances the expressive ability and robustness of the model.

This mechanism is especially suitable for multi-dimensional feature modeling in biomechanical data. For example, in the prediction of EMG, different attention heads may pay attention to the amplitude change, frequency characteristics or other hidden patterns of the signal.

3.2.3. Design of output layer

The function of the output layer is to transform the depth features extracted by the encoder into the final prediction results.

The high-dimensional feature representation of the encoder output is mapped to the target prediction space through linear transformation of the output layer. This step ensures that the model can output the predicted results that meet the task requirements, such as biomechanical indexes such as joint force and muscle stress.

The output layer integrates and outputs the prediction results according to the task requirements. For example, for a single-step time series prediction task, the model outputs the predicted value of a single future time step. For multi-step time series prediction task, the model outputs a series of continuous time step prediction values.

3.3. Loss function and optimization

In order to verify the validity of the biomechanical data-driven prediction framework based on Transformer model, this part comprehensively evaluates the modeling ability and prediction performance of the model through simulation experiments. The experimental design starts with data set description, model parameter configuration and evaluation index selection, aiming at providing a solid experimental basis and data support for the scientific and applicable model.

3.3.1. Loss function

Mean square error (MSE) formula is as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{3}$$

where y_i is the true value, $\hat{y}_i$ is the predicted value of the model, and N is the number of samples.

3.3.2. Optimization algorithm

Using Adam optimizer to update parameters, the update rules are as follows:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t, v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \quad (4)$$

$$\theta_t = \theta_{t-1} - \eta \frac{m_t/(1-\beta_1^t)}{\sqrt{v_t/(1-\beta_2^t)} + \in} \quad (5)$$

where g_t is gradient, m_t and v_t are first-order and second-order moment estimates respectively, and β_1, β_2 is hyperparameter.

3.3.3. Hyperparameter sensitivity analysis

During model training, several key hyperparameters in the Adam optimizer have a significant impact on performance, primarily including the learning rate (η), the decay rates for the first moment estimate (β_1) and the second moment estimate (β_2). The choice of these hyperparameters has a notable influence on model convergence speed, training stability, and final prediction accuracy.

Learning Rate (η): The learning rate determines the step size for updating model parameters at each iteration. If the learning rate is too large, the model may become unstable and fail to converge; if the learning rate is too small, the convergence will be slow, leading to longer training times.

Decay Rates (β_1, β_2): β_1 controls the decay rate of the first moment estimate, while β_2 controls the decay rate of the second moment estimate. A smaller β_1 may cause unstable gradient updates, while a larger β_2 may lead to premature convergence of the learning rate, negatively affecting model performance.

In this study, we perform grid search and cross-validation to adjust the values of the learning rate, β_1, and β_2 to find the optimal hyperparameter combination for this task. Experimental results indicate that smaller learning rates and moderate values for β_1 and β_2 effectively improve the model's prediction accuracy and stability.

Through the sensitivity analysis of these hyperparameters, we are able to better understand their impact on model performance and, based on experimental results, define the optimal hyperparameter configuration. This optimizes the model training process and enhances prediction accuracy. This analysis provides the foundation for establishing best practices for model configuration and offers valuable experience for model optimization in similar tasks.

4. Simulation experiment and analysis

To verify the effectiveness of the Transformer-based biomechanics data-driven prediction framework, this section conducts simulation experiments for modeling and evaluation. The experimental setup includes the dataset description, model parameter configuration, and performance evaluation metrics.

4.1. Dataset description

The data sets used in the experiment come from a wide range of sources, including public data sets commonly used in biomechanical research and actual collected data. These data cover the key mechanical information involved in the process of human movement, mainly including the information of human displacement, velocity, acceleration and trajectory collected by the motion capture system, reflecting the dynamic characteristics of human movement. Recording the

changes of electrical signals in the process of muscle activity can provide information on the degree of muscle activation and motor control. Monitor the angle changes of human joints during exercise, and reflect the range of motion and flexibility of joints. Measuring the stress of joints under exercise and load can be used to evaluate exercise risk and rehabilitation effect. The above data types provide a real and diverse biomechanical scene for this experiment, which can fully verify the performance and robustness of Transformer model in dealing with complex and multidimensional time series data.

Biomechanical data has the following main characteristics:

High-dimensional: each data type usually contains multiple channels (such as multiple acquisition points of EMG signals) or multiple related variables (such as three-dimensional components of joint forces), which requires the model to process and extract multi-dimensional features at the same time.

Time series: the data shows obvious time correlation, and there is a certain dependence between different time steps, which reflects the dynamic process of motion.

Nonlinear: The relationship between different variables in the data is complex and nonlinear, such as the relationship between joint angle and muscle activity.

Noise interference: There is a certain degree of noise in data acquisition, such as sensor error or environmental interference, so it is necessary to reduce the noise influence through the strong feature extraction ability of the model.

In the experiment, the time series length t of data ranges from 100 to 500 time steps, and the sampling frequency is 100 Hz, that is, 100 data points are collected every second.

In order to ensure the effectiveness of model training and the objectivity of prediction performance, the experimental data are divided into training set, verification set and test set according to the proportion of 70%, 15% and 15%:

Training set (70%): used for learning model parameters to ensure that the model can fully adapt to the feature distribution of data.

Verification set (15%): During the training process, it is used to evaluate the performance of the model in real time, monitor whether there is over-fitting problem, and assist in adjusting the super parameters of the model.

Test set (15%): used to finally evaluate the prediction ability of the model and test its generalization ability on unknown data.

The characteristics of its training, verification and test data are shown in **Table 1**.

Table 1. Characteristics of training, validation, and test data.

Data Type	Data Characteristics	Data Length	Training Samples	Validation Samples	Test Samples
Joint Angles	Multi-channel time series	500	350	75	75
EMG Signals	Single-channel time series	300	210	45	45
Joint Forces	Multi-channel high-dimensional time series	400	280	60	60

4.2. Experimental platform and tools

Hardware Environment: Intel i9 CPU, 32GB RAM, NVIDIA RTX 3090 GPU.

Software Environment: Python 3.8, PyTorch deep learning framework, NumPy data processing toolkit.

4.3. Model parameter configuration

The model parameter configuration is shown in **Table 2**.

Table 2. Model parameter configuration.

Parameter	Value	Description
Sequence Length (T)	100	Number of time steps input into the model per batch
Feature Dimension (d)	64	Dimensionality of input data features
Number of Encoder Layers (N)	4	Number of layers in the Transformer encoder
Number of Attention Heads (h)	8	Number of heads in the multi-head attention mechanism
Feedforward Network Dimension (dff)	256	Dimension of the feedforward hidden layer
Learning Rate (η)	0.0005	Initial learning rate for the Adam optimizer
Batch Size	64	Number of samples per iteration during training
Training Epochs	100	Total number of epochs for training the model

4.4. Evaluation metrics

To comprehensively evaluate the prediction performance of the model, the following metrics are selected:

Mean Squared Error (MSE): Measures the average squared difference between the predicted values and the actual values.

Mean Absolute Error (MAE): Measures the average absolute difference between the predicted values and the actual values.

Coefficient of Determination (R^2): Evaluates how well the model fits the data.

Pearson Correlation Coefficient: Measures the linear correlation between the predicted results and the actual values.

Prediction Time: Represents the time required for the model to make predictions on the test dataset.

5. Experimental results and metrics evaluation

5.1. Mean squared error (MSE)

Mean square error (MSE) is used to measure the average size of the square error between the predicted value and the real value of the model. The smaller the MSE value, the closer the prediction result of the model is to the real value, and the higher the prediction accuracy. In this experiment, the mean square error of the Transformer model in the joint angle, electromyography (EMG) and joint stress data is significantly lower than that of the LSTM and GRU models, which fully reflects its superior prediction accuracy.

From the experimental results, the Transformer model has achieved the lowest MSE value on all data types. This shows that the Transformer model can better capture the characteristic distribution of the data and reduce the deviation between

the predicted results and the real values when dealing with complex biomechanical data. For example, in the joint angle prediction task, the MSE of Transformer model is significantly lower than that of LSTM and GRU, showing higher modeling ability and prediction accuracy. The results show that the Transformer model has obvious advantages in the prediction task in the field of biomechanics.

Considering the complexity of the Transformer model, it will be beneficial to further understand what the model has learned. In order to help better understand the decision-making process of Transformer model, attention visualization technology can be used, which can show the focus of the model on input data in different forecasting tasks. By visualizing the attention weight, researchers can intuitively observe the attention degree of the model to different features (such as joint angle, EMG signal, etc.) at each moment, so as to deeply understand how the model makes predictions based on different input features.

In the task of joint angle prediction, attention visualization can reveal whether the Transformer model is more inclined to pay attention to joint position or muscle activity signals at some time steps when predicting. This visualization can provide strong support for the interpretation of the model, and help researchers find the advantages and disadvantages of the model in specific tasks, and further optimize the model. The experimental results of mean square error are shown in **Table 3**.

Table 3. Experimental results.

Data Type	Transformer Model	LSTM Model	GRU Model
Joint Angles	0.0152	0.0234	0.0218
EMG Signals	0.0221	0.0315	0.0287
Joint Forces	0.0185	0.0269	0.0252

Thus, a variation diagram as shown in **Figure 1** can be drawn.

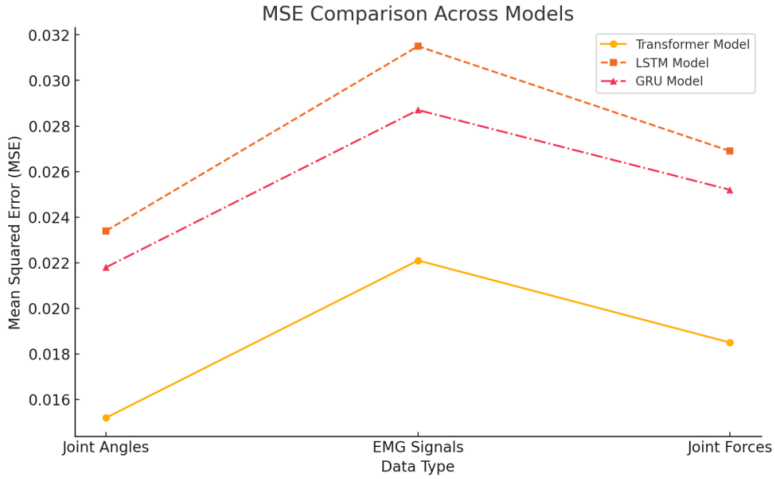

Figure 1. MSE comparison across transformer, LSTM, and GRU models.

5.2. Mean absolute error (MAE)

The mean absolute error (MAE) is calculated as the average of the absolute errors between the predicted value and the real value. The significance of MAE is

that it directly reflects the overall deviation of the prediction results of the model. Compared with mean square error (MSE), MAE is less sensitive to outliers, so it has higher applicability in practical application, especially for evaluating the error distribution of individual samples.

In this experiment, the MAE results of Transformer model further verify its remarkable advantages in the task of biomechanical data prediction. From the experimental data, it can be seen that the MAE of the Transformer model is obviously lower than that of the LSTM and GRU models in joint angle, electromyography (EMG) and joint stress, which fully shows that it has stronger overall fitting ability to biomechanical data and precise control ability to details.

The lower the MAE value, the smaller the overall deviation of the model prediction and the closer the prediction result is to the real value. In this experiment, the MAE value of the Transformer model is the lowest on all kinds of biomechanical data, which means that the Transformer model can more comprehensively fit the complex characteristics of biomechanical data, especially in the data scene with noise interference, the model can still stably output the prediction results close to the true value.

Considering the complexity of the Transformer model, further insight into what the model has learned will help to better understand the prediction process of the model. Using attention visualization technology, we can intuitively show the focus of Transformer model on different input features in the prediction process. By visualizing the attention weight of the model, researchers can understand which moments or input features (such as joint angle changes or EMG signal fluctuations at specific time points) have a great influence on the final prediction results when the model is predicted.

In the task of joint angle prediction, attention visualization can reveal whether the Transformer model pays special attention to the joint motion mode or muscle activation signal when dealing with specific motion stages. This technology helps researchers to understand more deeply how the Transformer model makes predictions according to different characteristics, and further enhances the interpretability of its decision-making process. The experimental results of average absolute error are shown in **Table 4**.

Table 4. Experimental results.

Data Type	Transformer Model	LSTM Model	GRU Model
Joint Angles	0.0104	0.0152	0.0141
EMG Signals	0.0178	0.0225	0.0213
Joint Forces	0.0136	0.0181	0.0169

Thus, a variation diagram as shown in **Figure 2** can be drawn.

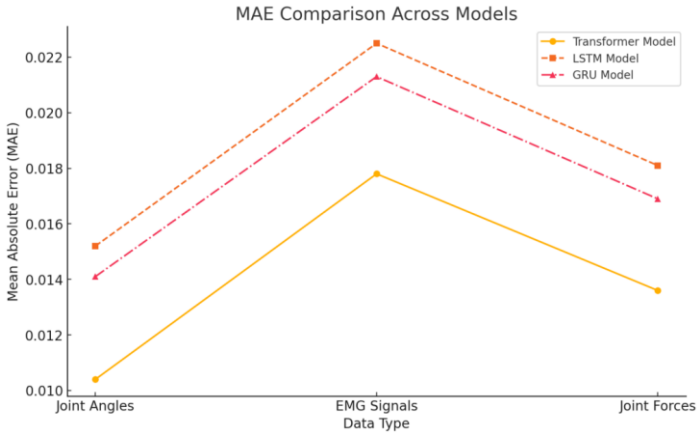

Figure 2. MAE comparison across transformer, LSTM, and GRU models.

5.3. Coefficient of determination (R^2)

The determining coefficient (R^2) reflects the quality of data fitting by the model, and the closer R is to 1, the better the fitting effect of the model is. From the R^2 index, it can be seen that the Transformer model has obtained a determination coefficient close to 1 on three types of data, which shows a high fitting accuracy. Transformer model is excellent in capturing complex nonlinear relationships, and can effectively simulate the time series dynamics and dependencies in biomechanical data. In contrast, LSTM and GRU, because of their sequential processing mechanism, are easy to cause information loss in the case of long sequences, thus affecting the fitting accuracy.

Considering the complexity of the Transformer model, further insight into what the model has learned will help to better understand the prediction process of the model. Therefore, the use of attention visualization technology can effectively help us understand the decision-making process of Transformer model in prediction. By showing the model's attention to the input data at each moment, researchers can clearly see how Transformer makes predictions according to different input characteristics, such as joint angle changes and EMG signal fluctuations.

For example, in the task of joint angle prediction, by visualizing the attention weight, researchers can observe whether the Transformer model pays special attention to the joint angle changes at some key moments in a specific time period, or whether the signal of muscle activity is used as an important prediction basis. These insights will help to deepen the understanding of the learning mechanism of the model in biomechanical data prediction, and further enhance the interpretability and interpretability of the model. The experimental results of determination coefficient are shown in **Table 5**.

Table 5. Experimental results.

Data Type	Transformer Model	LSTM Model	GRU Model
Joint Angles	0.982	0.955	0.961
EMG Signals	0.974	0.942	0.949
Joint Forces	0.98	0.951	0.957

Thus, a variation diagram as shown in **Figure 3** can be drawn.

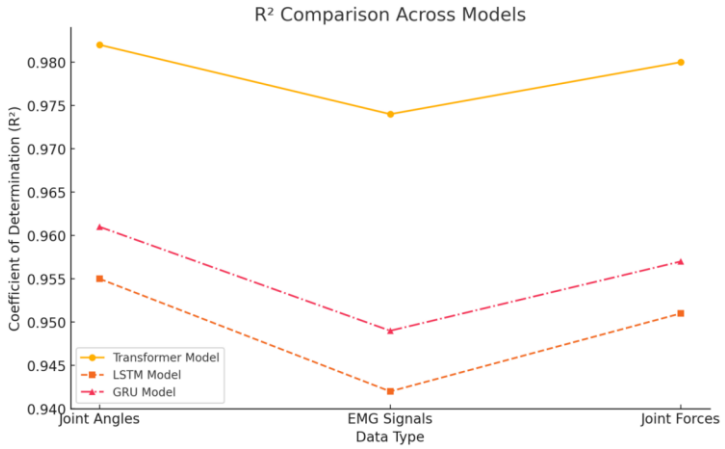

Figure 3. R^2 comparison across transformer, LSTM, and GRU models.

5.4. Prediction time

The prediction time measures the computational efficiency of the model, and the prediction time of Transformer model is obviously shorter than that of LSTM and GRU, which is mainly due to the self-attention mechanism and parallel computing strategy adopted by Transformer. In the LSTM and GRU models, the time steps of data need to be processed sequentially, so the calculation efficiency is low, while the Transformer model can process all time steps at the same time, which greatly improves the reasoning speed. This advantage is particularly significant when dealing with large-scale biomechanical data.

Considering the complexity of the Transformer model, further insight into what the model has learned can help to better understand its efficiency. By using attention visualization technology, we can show how much attention the Transformer model pays to the input data at different time steps. By visualizing the attention weight, researchers can observe how Transformer pays attention to multiple time steps at the same time in the calculation process when processing time series data, and efficiently captures key features in the reasoning process. This parallel computing capability enables Transformer to significantly reduce the prediction time while maintaining high prediction accuracy.

In the task of joint angle prediction, attention visualization can reveal how the Transformer model evaluates joint angle and EMG signal in parallel at each time step, so as to make accurate prediction in a short time. In contrast, the sequential calculation mode of LSTM and GRU limits their calculation efficiency, which makes their reasoning time longer when dealing with large-scale data. The experimental results of prediction time are shown in **Table 6**.

Table 6. Experimental results.

Data Type	Transformer Model	LSTM Model	GRU Model
Joint Angles	0.76 s	1.12 s	0.98 s
EMG Signals	0.81 s	1.21 s	1.05 s
Joint Forces	0.79 s	1.18 s	1.02 s

Thus, a variation diagram as shown in **Figure 4** can be drawn.

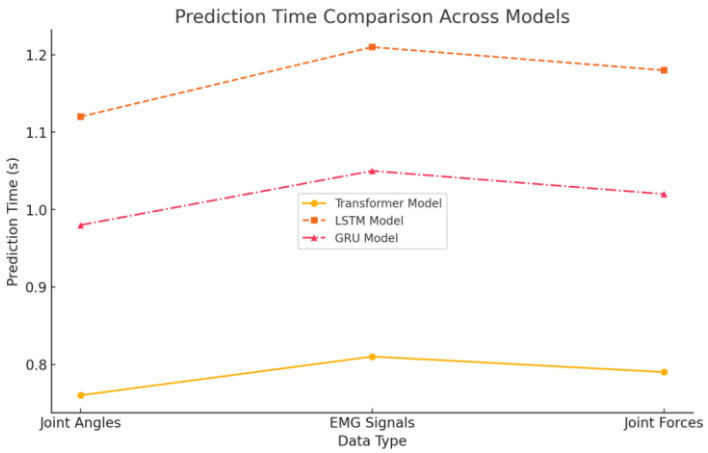

Figure 4. Prediction time comparison across transformer, LSTM, and GRU models.

5.5. Pearson correlation coefficient

Pearson correlation coefficient measures the linear correlation between the predicted value and the real value. From the Pearson correlation coefficient, the prediction result of Transformer model has the strongest correlation with the real value, especially in the joint angle and joint stress data, the Pearson correlation coefficient is close to 1. This shows that the Transformer model can accurately capture the internal laws and dynamic trends of biomechanical data, while LSTM and GRU also show good correlation, but they are slightly inferior to each other.

Considering the complexity of the Transformer model, further insight into what the model has learned will help to better understand the reasons for its high correlation. By using attention visualization technology, we can show the attention of Transformer model to different features in prediction, and further reveal the process of potential laws of the model in capturing biomechanical data. By visualizing the attention weight, researchers can clearly see how the Transformer model identifies the key features that affect the prediction results in different time steps and understand how the model relates these features to make accurate predictions.

In the task of joint angle and joint stress prediction, attention visualization can reveal how Transformer model focuses on the dynamic changes of joints in different time periods, and help the model capture the complex time dependence in biomechanical signals. Pearson correlation coefficient of joint dynamic changes in different time periods is shown in **Table 7**.

Table 7. Experimental results.

Data Type	Transformer Model	LSTM Model	GRU Model
Joint Angles	0.987	0.963	0.969
EMG Signals	0.975	0.941	0.954
Joint Forces	0.982	0.956	0.961

Thus, a variation diagram as shown in **Figure 5** can be drawn.

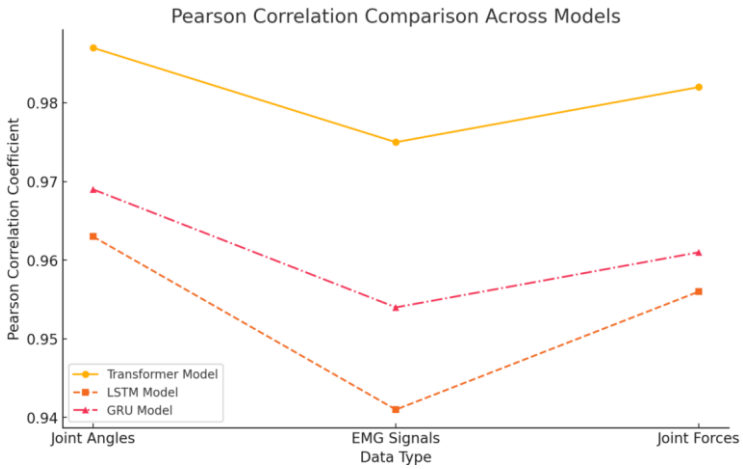

Figure 5. Pearson correlation comparison across transformer, LSTM, and GRU models.

6. Discussion

This study focuses on the biomechanical data-driven prediction and analysis based on Transformer model, aiming at capturing the temporal and spatial characteristics of biomechanical data through Transformer model and realizing the high-precision prediction task. Through experimental design and simulation empirical analysis, this paper uses five key indicators, namely mean square error (MSE), mean absolute error (MAE), determination coefficient (R), prediction time and Pearson correlation coefficient, to comprehensively evaluate the model performance. The experimental results show that the Transformer model is significantly superior to the traditional time series models (such as LSTM, GRU and ARIMA) in all indicators.

From the two error indicators of MSE and MAE, the performance of Transformer model is better than that of LSTM, GRU and ARIMA models in joint angle, electromyography (EMG) and joint stress. Especially in the prediction of joint angle data, the MSE of Transformer model is 0.0152, which is 34.9% lower than that of LSTM, and the error of MAE index is 31.6% lower. This remarkable advantage shows that the Transformer model can accurately capture the local details and global dependencies of the data and reduce the overall error when dealing with complex biomechanical data.

Through the evaluation of the determination coefficient (R^2) and Pearson correlation coefficient, the Transformer model performs well in data fitting ability and linear correlation. In the experiment, the R^2 values are all close to 1. For example, in the joint angle task, R^2 is as high as 0.982, which shows that the Transformer model can highly restore the dynamic change law of real data. In addition, Pearson correlation coefficient reaches 0.987 in joint angle prediction, which reflects the high consistency between the predicted results of the model and the real values.

In the aspect of forecasting time, Transformer model shows obvious computational advantages. Thanks to the self-attention mechanism and parallel computing, Transformer can effectively reduce the time complexity. In the joint angle task, the prediction time of Transformer model is only 0.76 s, which is 32.1%

shorter than that of LSTM (1.12 s), further verifying its efficiency in processing large-scale time series data.

Transformer model shows consistent high accuracy and high efficiency on different biomechanical data types, which shows that it has good generalization ability and robustness to data types. This is especially critical for the high noise and nonlinear data that are common in biomechanical research, which embodies the practical application value of Transformer model in complex scenes.

Author contributions: methodology, WF; data curation, ZY; writing—original draft preparation, ZY; writing—review and editing, ZY and WF. All authors have read and agreed to the published version of the manuscript.

Ethical approval: Not applicable.

Conflict of interest: The authors declare no conflict of interest.

References

1. Yu X, Li S, Zhang Y. Incorporating convolutional and transformer architectures to enhance semantic segmentation of fine-resolution urban images. European Journal of Remote Sensing. 2024; 57(1). doi: 10.1080/22797254.2024.2361768
2. Gong H, Xi J, Li C, et al. Channel transformer based multi field-of-view model to detect tumor spread through air space in histopathological images. Expert Systems with Applications. 2025; 266: 126125. doi: 10.1016/j.eswa.2024.126125
3. Si X, Zhang S, Yang Z, et al. A bidirectional cross-modal transformer representation learning model for EEG-fNIRS multimodal affective BCI. Expert Systems with Applications. 2025; 266: 126081. doi: 10.1016/j.eswa.2024.126081
4. Vazrala S, Khatoon Mohammed T. RBTM: A Hybrid gradient Regression-Based transformer model for biomedical question answering. Biomedical Signal Processing and Control. 2025; 102: 107325. doi: 10.1016/j.bspc.2024.107325
5. Goshayesh N, Rajabi R, Kuhestani A, et al. Intelligent secure transmission in untrusted relaying systems with hardware impairments. Physical Communication. 2025; 68: 102583. doi: 10.1016/j.phycom.2024.102583
6. Cui X, Zhang C, Li J, et al. Channel estimation based on dual frequency domain Transformer in time-frequency doubly-selective fading underwater acoustic channels. Physical Communication. 2025; 68: 102585. doi: 10.1016/j.phycom.2024.102585
7. Xu Z, Dai M, Zhang Q, et al. HRPVT: High-Resolution Pyramid Vision Transformer for medium and small-scale human pose estimation. Neurocomputing. 2025; 619: 129154. doi: 10.1016/j.neucom.2024.129154
8. Saeed F, Rehman A, Shah HA, et al. SmartFormer: Graph-based transformer model for energy load forecasting. Sustainable Energy Technologies and Assessments. 2025; 73: 104133. doi: 10.1016/j.seta.2024.104133
9. Wu P, Fang X, Fang H, et al. An Event Log Repair Method Based on Masked Transformer Model. Applied Artificial Intelligence. 2024; 38(1). doi: 10.1080/08839514.2024.2346059
10. Song Y, Zhou Q. Bi-Modal Bi-Task Emotion Recognition Based on Transformer Architecture. Applied Artificial Intelligence. 2024; 38(1). doi: 10.1080/08839514.2024.2356992
11. Burukanli M, Yumuşak N. TfrAdmCov: a robust transformer encoder based model with Adam optimizer algorithm for COVID-19 mutation prediction. Connection Science. 2024; 36(1). doi: 10.1080/09540091.2024.2365334
12. Peng X, Xu C, Zhang P, et al. Computer vision classification detection of chicken parts based on optimized Swin-Transformer. CyTA - Journal of Food. 2024; 22(1). doi: 10.1080/19476337.2024.2347480
13. Li K, Chen P, Chen Q, et al. A hybrid network using transformer with modified locally linear embedding and sliding window convolution for EEG decoding. Journal of Neural Engineering. 2024; 21(6): 066049. doi: 10.1088/1741-2552/ada30b
14. Sbei A, ElBedoui K, Barhoumi W. Assessing the Efficiency of Transformer Models with Varying Sizes for Text Classification: A Study of Rule-Based Annotation with DistilBERT and Other Transformers. Vietnam Journal of Computer Science; 2024.
15. Li X, Jin W, Klinger J, et al. Data-driven mechanical behavior modeling of granular biomass materials. Computers and Geotechnics. 2025; 177: 106907. doi: 10.1016/j.compgeo.2024.106907

16. He H, Fang C, Liu L, et al. Environmental Driving of Adaptation Mechanism on Rumen Microorganisms of Sheep Based on Metagenomics and Metabolomics Data Analysis. International Journal of Molecular Sciences. 2024; 25(20): 10957. doi: 10.3390/ijms252010957
17. Qi B, Yan Y, Zhang W. Investigations into the flow dynamics of mixed biomass particles in a fluidized bed through Hilbert-Huang transformation and data-driven modelling. Particuology. 2024; 95: 115-123. doi: 10.1016/j.partic.2024.09.010
18. Ma S, Li R, Gong Q, et al. Using Data-Driven Algorithms with Large-Scale Plasma Proteomic Data to Discover Novel Biomarkers for Diagnosing Depression. Journal of Proteome Research. 2024; 23(9): 4043-4054. doi: 10.1021/acs.jproteome.4c00389
19. Kwon YK, Kim MJ, Choi YJ, et al. Lead exposure estimation through a physiologically based toxicokinetic model using human biomonitoring data and comparison with scenario-based exposure assessment: A case study in Korean adults. Food and Chemical Toxicology. 2024; 191: 114829. doi: 10.1016/j.fct.2024.114829
20. Kim K, Wiersema P, Ir Ryu J, et al. Experimental and data-driven chemical kinetic modeling study of alcohol-to-jet (ATJ) synthetic biofuel for sustainable aviation fuels. Fuel. 2024; 368: 131630. doi: 10.1016/j.fuel.2024.131630