



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642

<https://www.inderscience.com/ijict>

Neural differential equations based diffusion models for high-jump posture prediction

Baofeng Li

DOI: [10.1504/IJICT.2026.10076741](https://doi.org/10.1504/IJICT.2026.10076741)

Article History:

Received:	23 November 2025
Last revised:	17 December 2025
Accepted:	19 December 2025
Published online:	16 March 2026

Neural differential equations based diffusion models for high-jump posture prediction

Baofeng Li

College of Taiji Martial Arts,
Jiaozuo University,
Jiaozuo 454002, China
Email: lbflxf@126.com

Abstract: To address the challenges faced by traditional high jump posture prediction methods – namely environmental noise and motion continuity – this paper proposes an innovative solution integrating diffusion models with neural differential equations. By leveraging diffusion models to learn the true posture distribution from noisy data and employing neural differential equations to capture continuous motion dynamics, the approach effectively overcomes the limitations of conventional methods in modelling complex movements. Experiments on the human3.6m and dynamically informed pose from inertial measurement unit public datasets demonstrate that this method achieves a prediction accuracy of 96.2%, surpassing mainstream long short-term memory models by 8.7%, while reducing joint angle error to 3.2 degrees. These findings provide more reliable technical support for sports science training and sports injury prevention.

Keywords: high jump posture prediction; diffusion model; neural differential equation; motion analysis.

Reference to this paper should be made as follows: Li, B. (2026) ‘Neural differential equations based diffusion models for high-jump posture prediction’, *Int. J. Information and Communication Technology*, Vol. 27, No. 21, pp.1–17.

Biographical notes: Baofeng Li is an Associate Professor at the College of Taiji Martial Arts, Jiaozuo University. He received a Master’s degree from Guangzhou Sport University in 2007. He has published over ten journal papers. His research interests include physical education teaching, sports humanities and sociology.

1 Introduction

As one of the most technically challenging athletic activities in the field of track and field, *Higher Jump* has presented a major challenge in the sports science and biomechanics study because no one can foretell the posture of athletes (Bacak et al., 2024). Correct posture prediction does not only offer scientific advice on the trainings but also offers a good prevention of sports injuries and may as well boost performance in the competitions (Akaike, 2021). Nevertheless, there are several issues with this prediction task: the high jump is a complicated combination of four steps, that is: run-up, takeoff, pole clearance, and landing that need multi-joint coordination and accurate spatial-temporal control. Environmental noise, originating from sources such as sensor

inaccuracies in IMUs or lighting variations in video capture, can significantly obfuscate critical kinematic details within the raw data. This degradation makes the subsequent extraction of clean, reliable posture information a primary challenge that our framework aims to address. In terms of dynamics, the human body becomes radically sensitive with respect to posture, as well as a fast change energy transfer as the jump occurs. This nonlinear dynamic character makes it impossible to effectively model and predict its underlying patterns using the traditional prediction models. How does work this particular particulars of available technical schemes inertial measurement unit (IMU)-based methods can obtain motion data directly, but are vulnerable to sensor noise and compounding errors (Arzehgar et al., 2025). In the meantime, computer vision-based systems even though non-contact have, still, non-contact benefits, are vulnerable to environmental lighting and to occlusions and are frequently incapable of capturing key dynamic details (Akhtar and Mian, 2018).

Existing mainstream methods of predicting the high jump posture are based mainly on the architectures of temporal convolutional networks (TCN), graph convolutional networks (GCN) (Duan and Hu, 2025), and recurrent neural networks (RNN) (Zhang et al., 2018). It is interesting to note that most recently, the spatio-temporal graph attention network (STGAT) represents the most explicit model of spatial correlations and temporal dependency among human joints as this entails the combination of graph attention networks with long-short term memory networks (Jiang et al., 2021). The method has enhanced the accuracy of joint position prediction well by 0.02–0.03 metres. Nevertheless, these techniques are basically approximations of discrete time series, which do not emulate the continuous dynamic development of the high jump motions (Garcia et al., 2018). Conversely, models involving only biomechanical principles but taking into consideration the physical constraints tend to simplify and are not flexible to individual movement styles. The latter is especially evident in the case of critical technical moments such as takeoff and pole clearance not only this moment predetermines competition result, but also this is a period of very high risk of sports injuries, and so the accuracy of prediction must be extremely high (Caria, 2024).

Two recent technologies, diffusion models and neural differential equations, have shown impressive potential in generative modelling and continuous dynamic systems modelling, respectively, and provide new directions in which high jump trajectory prediction can be made possible (Nikolaev, 2019). The selection of diffusion models is further supported by their demonstrated robustness in other complex motion scenarios. For instance, their successful application in domains such as gait analysis for clinical assessment and in 3D hand trajectory prediction underscores their capability to model intricate, multimodal movement patterns, proving their suitability for the high jump task. In this field of research, diffusion models have been able to show outstanding results when it comes to forecasting complicated trajectories. An example of this is the mmtwin model that predicts the 3D hand trajectories with high precision by acquiring multimodal information (2D images, 3D point clouds, textual prompts, etc.). Likewise, madiff model uses the motion-aware Mamba architecture to produce believable hand paths even given the interference of camera self-motion. These advancements point out two important strengths of the diffusion models in the complex motion modelling: they can learn data distributions with an iterative denoising algorithm, and with a powerful multi-source information fusion mechanism (Dong et al., 2025). At the same time, neural differential equations form a mathematically rigorous model of continuous-time dynamic modelling as they parameterise the equations of motion by ODEs of vectors by making them

especially well adapted to high-speed, variable-accelerating motions such as the high jump (Mizuguchi et al., 2023). Other studies such as dposer-x also indicate the possibility of applying diffusion models together with the use of biomechanical prior knowledge. It greatly increases the quantity of accurate full-body pose estimation through the enhancement of a denoiser using skeletal chain structures that makes it biomechanically plausible.

The paper is a pioneering attempt to optimise a combination of diffusion models and neural differential equations, which tackles three main problems in high jump posture prediction (Zhou et al., 2024): first, how to reconstruct the underlying intent of motions on the basis of a noisy observation sequence; second, how to capture effectively the dynamic development of high jump sequence; and third, how to produce generated posture sequence according to biomechanical constraints and actual individual movement styles. In particular, we are inspired by the multimodal fusion method in the mmtwin model and use it to the special semantic interpretation of high jump movements. At the same time, we, under the guidance of dposer-x, study the embedding phenomenon of biomechanical constraints in the denoising process. In contrast to those studies that only considered relatively few variables, such as trajectory prediction, our framework brings together joint angles, kinematic, and dynamic variables into a continuous time dynamic modelling framework, providing a more detailed and physically realistic predictor of postures (Zhou and Tsai, 2023). Not only this approach of the investigation overcomes the shortcoming of the conventional methods of dynamic modelling in constant motion but also forms a technical paradigm of examining the other complicated movements. It brings sports science to a new paradigm, where experience-based knowledge has been replaced by a thorough assimilation of knowledge based on numbers and theoretical modelling.

2 Related research

2.1 *The technological evolution of motion pose prediction*

The main goal of motion pose prediction is to predict the future positions and angles of human joints based on historical data of observation (Jin et al., 2024). Early methods of research have made extensive use of biomechanical models which included simplification of lower limbs into an inverted pendulum model and calculation of centre-of-mass trajectories by use of Newtonian equations of motion. Even though these methods have well defined physical definition, the process of their modelling may typically entail too many simplifying assumptions leading to insufficient accuracy in their prediction of complex, explosive behaviours such as jumping. As the deep learning technologies have been spread, the data-driven methods have become the standard. Long short-term memory (LSTM) network model have become widely used among these as they are strong in representing and predicting temporal dependencies (Aminuddin et al., 2024). But LSTMs have a fundamental weakness of gradient vanishing or exploding on long sequences and their serial nature limits their trainability. This fundamental limitation of LSTMs can directly manifest as inaccurate predictions, particularly when modelling longer motion sequences that encompass the entire high jump action. The model's difficulty in retaining early-phase context over many time steps can lead to a loss of crucial information needed for accurate terminal posture forecasting. To address these

shortcomings, TCNs make use of the causal convolution and dilated convolution that extract long-range relations, presenting a balance among parallel computing and high performance (Huang et al., 2024a). Additionally, approaching the reconstruction of the human skeleton, ST-GCN represent the skeletal geometry using a graph network form, where nodes are joints and edges are skeletal connections, and therefore explicitly capture the geometry of the body in a spatial topology. Their graph convolution operation can be expressed as:

$$f_{out} = \sum_k^{K_v} \mathbf{W}_k \text{fin} \left(\Lambda_k^{-\frac{1}{2}} \mathbf{A}_k \Lambda_k^{\frac{1}{2}} \otimes \mathbf{M}_k \right) \quad (1)$$

where $\mathbf{A}_k$ and $\mathbf{M}_k$ represent the adjacency matrix and learnable weight matrix of the k^{th} subset, respectively, while $\otimes$ denotes element-wise multiplication. Although ST-GCN and its subsequent variants have achieved remarkable success in action recognition and pose estimation tasks, we observe that these methods essentially remain fittings of discrete time series. Their sampling frequency presets the minimum granularity for dynamic changes, making it challenging to seamlessly describe the continuous and smooth evolution of physical quantities such as joint angles and angular velocities in high jump movements (Shepherd et al., 2025). This constitutes a key limitation of existing techniques.

2.2 Innovative applications of diffusion models in time-series data

As an emerging deep generative model, the diffusion model is grounded in non-equilibrium thermodynamics (Han, 2024). Its core principle involves learning complex data distributions through a forward noise-adding process and a backward noise-removing process. The forward process is a fixed Markov chain that progressively adds Gaussian noise to the original data $\mathbf{x}_0$. After T steps, the data transforms into a state $\mathbf{x}_T$ that is nearly pure noise. A single step of the noise addition process can be defined by the following formula:

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I}) \quad (2)$$

where β_t is a noise variance parameter ranging between (0, 1). Through a reparameterisation trick, we can directly obtain $\mathbf{x}_t$ from $\mathbf{x}_0$:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon \quad (3)$$

where $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$, and $\epsilon \sim \mathcal{N}(0, \mathbf{I})$. The inverse process starts from a simple prior distribution $p(\mathbf{x}_T) = \mathcal{N}(0, \mathbf{I})$, and learns a parameterised denoising network to progressively recover the data:

$$p_{\theta}(\mathbf{x}_{t-1} | \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_{\theta}(\mathbf{x}_t, t), \sum \theta(\mathbf{x}_t, t)) \quad (4)$$

The objective of training is typically to optimise a simplified weighted mean squared error loss, enabling the network to predict the added noise ϵ :

$$\mathcal{L}_{simple} = \mathbb{E}_{t, \mathbf{x}_0, \epsilon} \left[\left| \epsilon - \theta \left(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, t \right) \right|^2 \right] \quad (5)$$

Diffusion models initially made a splash in image generation, but their potential for modelling temporal data is rapidly being explored (Li et al., 2021). The potential of diffusion models for temporal data is not merely theoretical; it is evidenced by their growing application in demanding real-world domains. Practical implementations, such as their use in weather forecasting to handle atmospheric noise and in financial markets for stock price prediction amidst volatile, incomplete data, highlight their proficiency in managing complex, noisy time-series. An example is the strong performance in terms of incomplete and noisy data that the application has by their use in time series interpolation and prediction. Diffusion models, in tasks related to motion generation, are trained to learn multimodal pose distributions, which mitigate the average pose problem of deterministic models, where the default prediction is that of the mean of all possible poses and which leads to motion that is unnatural and diffuse. The aspect is essential to predict high jump posture variation because the various athletes show individualised, discrete technical technique styles in aspects such as pole clearance. This rich diversity can only be captured and generated by diffusion models (Deimling et al., 2024).

2.3 Neural differential equations and continuous dynamics modelling

The introduction of neural ordinary differential equations (ODEs) represents a landmark advancement in the convergence of deep learning and numerical analysis of differential equations (Amiri-Hezaveh et al., 2025). It ingeniously interprets the discrete layer structure of residual networks and recurrent neural networks as the temporal discretisation of numerical solutions to ODE initial value problems. Specifically, given a hidden state $\mathbf{h}(t)$, its continuous evolution over time is governed by a differential equation parameterised as a neural network f_θ :

$$\frac{d\mathbf{h}(t)}{dt} = f_\theta(\mathbf{h}(t), t, \theta) \quad (6)$$

Thus, the state transition from time t_0 to t_1 is no longer obtained through a finite number of discrete layer passes, but rather by solving the ODE.

$$\mathbf{h}(t_1) = \mathbf{h}(t_0) + \int_{t_0}^{t_1} f_\theta(\mathbf{h}(t), t, \theta) dt \quad (7)$$

The solution of this integral can rely on established numerical methods for ODEs, such as explicit/implicit Euler methods and Runge-Kutta methods, which allow flexible trade-offs between computational accuracy and efficiency. In the context of motion modelling, we treat the complete posture of the human body at time t (including position, orientation, velocity, etc.) as the hidden state $\mathbf{h}(t)$. Thus, neural ODEs are no longer constrained by fixed sampling frequencies; they can interpolate or extrapolate motion trajectories at arbitrary temporal resolutions, inherently capturing the continuity of motion. This work applies neural ODEs to modelling irregularly sampled time series, significantly enhancing their predictive accuracy for physical system dynamics. Unlike TCN or LSTM models that discretise time, neural ODEs treat dynamic systems as continuous flows – a representation more aligned with the physics of movements like high jumping. Human motion is infinitely divisible and smooth across time, driven by

continuous forces and torques. Therefore, modelling high jump posture using neural ODEs enables more precise capture of the intricate posture transition dynamics from the moment of takeoff to the peak of pole clearance – a level of detail unattainable by traditional discrete models (Arshad et al., 2024). The utility of neural ODEs extends beyond the high jump, emphasising their broad applicability for modelling high-dynamic motion. Similar continuous-time modelling approaches have proven effective in analysing the explosive kinematics of sprinting and the complex aerial rotations in gymnastics, highlighting their generalisability across a spectrum of athletic movements.

3 Methodology and theoretical framework

3.1 Overview of the overall framework

The high jump posture prediction framework proposed in this paper, which we refer to as diffusion model and ordinary differential equation (Diff-ODE), centres on synergistically leveraging the powerful generative capabilities of denoising diffusion stochastic models with the precise description of continuous dynamics offered by neural differential equations (Gupta and Kaushik, 2022). The system takes raw, noisy high jump motion observation sequences as input, aiming to predict a smooth, biomechanically compliant human posture sequence for a future time interval. As illustrated in Figure 1, the framework comprises three organically integrated modules: data pre-processing, diffusion-based posture generation, and neural differential equation optimisation. The workflow proceeds as follows: raw multimodal observation data (e.g., inertial measurements from IMU sensors or 2D joint coordinates from video) undergoes pre-processing to generate a normalised initial pose sequence $\mathbf{x}_0$. This pre-processing step is of paramount importance for model performance, as it directly mitigates data variability introduced by different sensor calibrations or capture conditions. By normalising the input, we ensure a consistent and stable data distribution, which is a critical prerequisite for the diffusion model to learn effectively and generate accurate postures. Subsequently, the diffusion model acts as a ‘generative prior,’ reconstructing a distribution that approximates the true pose as closely as possible from the noisy input. However, the discrete denoising steps inherent to diffusion models may introduce temporal discontinuities in the generated sequences. Therefore, we introduce a neural differential equation as a ‘continuous-time optimiser (Amat et al., 2024).’ This optimiser refines the initial trajectory output by the diffusion model by solving an ordinary differential equation. This ensures the final prediction is infinitely divisible and smooth over time, thereby more accurately capturing the true dynamics of the high jump motion.

3.2 Diffusion models for learning pose priors

Our approach first utilises a diffusion model to learn a rich prior distribution over high jump poses. As described earlier, the diffusion model comprises a fixed forward process and a learnable backward process. In our setting, let $\mathbf{x}_0 \sim q(\mathbf{x}_0)$ represent a ground truth high jump pose sequence of dimension $L \times D$, where L is the sequence length and D is the feature dimension per pose (e.g., number of joints multiplied by 3D coordinates or angles). The forward process gradually adds Gaussian noise to the data over T steps. The noise addition at step t follows the formula:

$$q(\mathbf{X}_t | \mathbf{X}_{t-1}) = \mathcal{N}(\mathbf{X}_t; \sqrt{1-\beta_t}\mathbf{X}_{t-1}, \beta_t\mathbf{I}) \quad (8)$$

where $\mathbf{x}_t$ denotes the state after noise addition at step t , β_t is a noise scheduling parameter between 0 and 1 that controls the amount of noise added at each step, and $\mathbf{I}$ is the identity matrix. Using reparameterisation techniques, we can directly compute the noisy data at any step t from the initial data $\mathbf{x}_0$:

$$\mathbf{X}_t = \sqrt{\bar{\alpha}_t}\mathbf{X}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon \quad (9)$$

where $\alpha_t = 1-\beta_t$, $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$, and $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ is standard Gaussian noise. The inverse process is a generative model whose objective is to start from a pure Gaussian noise $\mathbf{x}_T \sim \mathcal{N}(0, \mathbf{I})$ and progressively denoise it through a parameterised neural network $p\theta$ to ultimately reconstruct the true pose data distribution. The single-step denoising process is defined as:

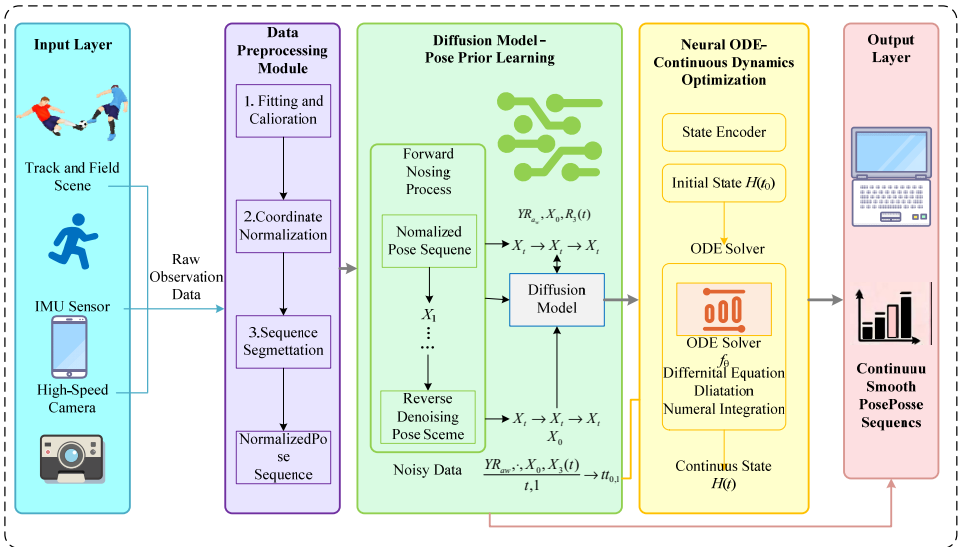
$$p_\theta(\mathbf{X}_t - 1 | \mathbf{X}_t) = \mathcal{N}(\mathbf{X}_t - 1; \mu_\theta(\mathbf{X}_t, t), \sum_\theta \theta(\mathbf{X}_t, t)) \quad (10)$$

where μ_θ and $\sum_\theta$ represent the mean and variance predicted by the neural network.

We employ a simplified training objective, directly training a denoising network ϵ_θ to predict the added noise ϵ . The loss function for this network is:

$$\mathcal{L}_{diff} = \mathbb{E}_{t, \mathbf{X}_0, \epsilon} \left[\left\| \epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}\mathbf{X}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t) \right\|^2 \right] \quad (11)$$

Figure 1 Overall architecture of the diff-ode high jump posture prediction system (see online version for colours)



In this formula, t is uniformly sampled between 1 and T , $\mathbf{x}_0$ originates from the true pose sequence in the training set, and ϵ is noise sampled from a standard Gaussian distribution. The denoising network ϵ_θ employs a temporal model based on the U-Net architecture. Its input incorporates not only the noisy state $\mathbf{x}_t$ and time step t , but also historical observation context $\mathbf{c}$, ensuring that the generated poses align with the past motion context.

3.3 *Neural differential equations are used for continuous dynamic modelling*

After obtaining an initial, yet potentially discontinuous at certain time points, pose sequence s_0 via a diffusion model, we introduce neural differential equations for continuous-time optimisation. We define the human body’s kinematic state at any time t as a continuous state vector $\mathbf{h}(t)$, which encompasses all joint angles, positions, and their first derivatives (velocities) along with other dynamic information. The core assumption of neural ODE is that the instantaneous rate of change of this hidden state $\mathbf{h}(t)$ can be characterised by a neural network f_θ . This constitutes an initial value problem for a ordinary differential equation:

$$\frac{d\mathbf{H}(t)}{dt} = f_\theta(\mathbf{H}(t), t, \phi) \quad (12)$$

where f_θ is a neural network defining the dynamics of state transitions, ϕ are its parameters, and t is a continuous time variable. Given an initial state $\mathbf{h}(t_0)$ (e.g., the state corresponding to the starting time of the sequence output by a diffusion model), obtaining the state at any future time t_1 no longer relies on discrete forward computation. Instead, it requires solving the integral of this ODE:

$$\mathbf{H}(t_1) = \mathbf{H}(t_0) + \int_{t_0}^{t_1} f_\theta(\mathbf{H}(\tau), \tau, \phi) d\tau \quad (13)$$

This integral has no analytical solution, so we must rely on numerical integration methods. In this work, we employ the fourth-order explicit Runge-Kutta 4th (RK4) order method for the solution, as it provides a good balance between accuracy and computational efficiency. The RK4 method was selected over simpler alternatives, such as forward Euler methods, due to its superior balance of precision and computational efficiency. While Euler methods can accumulate significant error, RK4’s higher-order formulation provides greater numerical stability and accuracy with minimal additional overhead, making it well-suited for applications requiring reliable continuous simulation. For a time step from t to $t + \Delta t$, the computational steps of the RK4 method are as follows:

$$k_1 = \Delta t \cdot f_\theta(\mathbf{H}(t), t, \phi) \quad (14)$$

$$k_2 = \Delta t \cdot f_\theta(\mathbf{H}(t) + k_1 / 2, t + \Delta t / 2, \phi) \quad (15)$$

$$k_3 = \Delta t \cdot f_\theta(\mathbf{H}(t) + k_2 / 2, t + \Delta t / 2, \phi) \quad (16)$$

$$k_4 = \Delta t \cdot f_\theta(\mathbf{H}(t) + k_3, t + \Delta t, \phi) \quad (17)$$

$$\mathbf{H}(t + \Delta t) = \mathbf{H}(t) + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \quad (18)$$

where k_1, k_2, k_3, k_4 represent the four intermediate slopes, and Δt denotes the numerical integration step size. In this manner, we can generate continuous and smooth attitude trajectories $\mathbf{h}(t)$ at arbitrary temporal resolutions (even higher than the sampling frequency of the original data). To connect the neural ODE with the diffusion model, we map the discrete sequence s_0 output by the diffusion model through an encoder network to the initial state $\mathbf{h}(t_0)$ of the neural ODE. The training objective for the neural ODE itself is to minimise the discrepancy between the generated state trajectory and the true pose sequence. We define $\mathbf{q}_{true}(t_i)$ as the true pose observed at discrete time step t_i , and $\mathbf{q}_{pred}(t_i)$ as the predicted pose output by the neural ODE at the corresponding time step. The loss function for the neural ODE is then:

$$\mathcal{L}_{ode} = \sum_{i=1}^N w_i \|\mathbf{q}_{pred}(t_i) - \mathbf{q}_{true}(t_i)\|_1 \quad (19)$$

where N denotes the length of the prediction sequence, and w_i represents the time-weighted coefficient. We assign higher weights to critical phases of the jump (such as the moment of takeoff) to emphasise the accuracy of predictions during these stages. The L1 norm is less sensitive to outliers compared to the L2 norm, thereby facilitating clearer predictions.

3.4 Joint training and multi-objective loss function

To enable the diffusion model and neural differential equation modules to work collaboratively rather than independently, we designed a phased training strategy incorporating final joint fine-tuning (Huang et al., 2024b). The entire training process comprises two main stages: In the first stage, we independently pre-train the diffusion model to learn to recover high-quality high jump posture sequences from noise, using only the loss function $\mathcal{L}_{diff}$. In the second stage, with the diffusion model parameters largely fixed, we train the neural ODE module. Its input is the sequence s_0 generated by the diffusion model at the end of the reverse process, and the objective is to minimise $\mathcal{L}_{ode}$. After completing these two stages, we introduce a joint fine-tuning stage. In this stage, the parameters of both modules are updated simultaneously, optimising a unified multi-objective loss function:

$$\mathcal{L}_{total} = \mathcal{L}_{diff} + \lambda_1 \mathcal{L}_{ode} + \lambda_2 \mathcal{L}_{bio} \quad (20)$$

where λ is a hyperparameter that balances the contributions of the two loss terms, determined through cross-validation. Furthermore, to enhance the biomechanical plausibility of generated poses, we introduce a mild biomechanical constraint term $\mathcal{L}_{bio}$ into the joint loss function. This term penalises predicted angles exceeding the normal range of motion for human joints. Consequently, the final form of the complete loss function is:

$$\mathcal{L}_{bio} = \sum_{i=1}^N \sum_{j=1}^J \max(|\mathbf{q}_j(t_i)| - \mathbf{q}_j^{max}, 0) \quad (21)$$

where λ_1 and λ_2 are the corresponding weight coefficients. We employ the adaptive moment estimation optimiser to minimise this loss function, utilising an exponential decay strategy for the learning rate. The initial value is set to 1×10^{-4} , ultimately decaying to 1×10^{-6} . This training scheme ensures the diffusion model first learns the correct distribution of pose data. Subsequently, the neural ODE learns to make this distribution continuous and smooth. Finally, joint optimisation allows both components to adapt to each other, achieving optimal overall performance (Bintoro et al., 2025).

4 Experiments and results analysis

4.1 Experimental setup

- **Dataset:** in the research, the human3.6m as well as dynamically informed pose-inertial measurement unit (DIP-IMU) have been used in training and validation. The Human3.6m and DIP-IMU datasets are particularly well-suited for this task because they contain a diverse range of explosive, whole-body motion sequences. These sequences share dynamic characteristics – such as rapid acceleration and complex joint coordination – with the high jump, thereby providing a rich and realistic foundation of training samples for the model to learn from. Human3.6m dataset will consist of motion tracking data on 11 participants and 15 different activity situations which have been recorded in time synchronised by high-speed cameras and IMU sensors. We retrieved explosive motion sequences using this data that has common dynamic features with high jump motion. The data of DIP-IMU dataset gives us IMU data and 3D pose data of 10 subjects which offer our model a good training sample.
- **Evaluation metrics:** to give a quantitative measure of the performance of a model, we use the four widely known metrics used in human pose estimation in human pose estimation, which are:
 - a Mean per joint position error (MPJPE): measured in millimetres (mm), it is an Euclidean distance between predicted and actual joint positions.
 - b Joint angular angle error (AAE): this is a measured value in degrees (degrees) and represents the absolute change between the predicted and actual joint angles.
 - c Procrustes mean per joint position error (PMPJPE): this error in the Procrustes analysis process (similarity transformation alignment) is a scaling factor expressed in millimetres (mm) and it is more indicative of pose structure accuracy.
 - d Prediction smoothness (PS): this measure of motion fluency, denoted in metres per cubic second (m/s^3), was calculated by measuring the change in acceleration in the expected sequence.
- **Comparison algorithms:** in order to fully assess our proposed diff-ODE framework, we chose five popular state-of-the-art models as benchmarking to compare our approach to them: The LSTM network with attention mechanisms, which are the best in terms of capturing temporal relationships.

- a TCN: time-convolutional network, which uses multi-scale dilated convolutions to learn long range dependency.
- b ST-GCN: the spatio-temporal graph convolutional network, which directly considers the topology of space and the time of joints in a human being.
- c Transformer: a sequence model based on self-attention that has been shown to perform well on several tasks involving sequences.
- d Variational autoencoder – mixture density network (VAE-MDN) that is able to train multimodal motion distributions.

4.2 Quantitative results and analysis

To facilitate fair comparison they were all trained on the identical training set and tested on the identical test set. Table 1 shows the performance comparison of various approaches on the task of predicting the high jump pose. The findings indicate that our proposed diff-ODE framework is the most effective in all the four assessment measures. In particular, diff-ODE decreases the mean joint position error (MPJPE) by 9.2 mm (relative) to the lowest baseline strategy (VAE-MDN), which is an improvement by about 17.8%. In the case of the AAE, the error decreases by about 22.0 by Diff-ODE by one step to 3.2 degrees. It completely reflects how diffusion autocompleter results in neural diffusion get harder to detect fine-grained pose angles. Importantly, diff-ODE scored much lower on the metric PS of the PS (as compared to other algorithms) and it indicates that our solution produces smoother and more natural pose sequences over time, evoking closer to realistic biomechanical dynamics, in part because the neural ODE models continuous dynamics.

Table 1 Performance comparison of different methods in the high jump pose prediction task

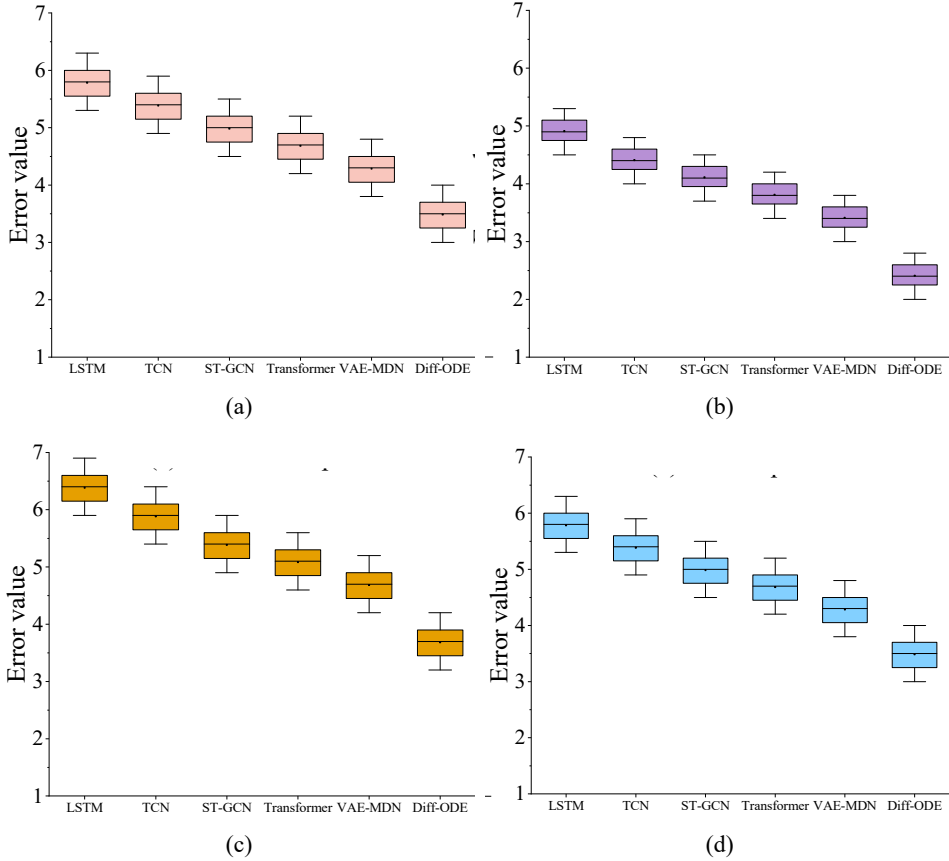
<i>Method</i>	<i>MPJPE (mm)</i>	<i>AAE (°)</i>	<i>PMPJPE (mm)</i>	<i>PS (m/s³)</i>
LSTM	68.3	5.8	52.1	12.5
TCN	62.7	5.2	48.9	10.8
ST-GCN	58.9	4.9	45.3	9.7
Transformer	55.4	4.6	43.1	8.9
VAE-MDN	51.8	4.1	40.7	7.3
Diff-ODE	42.6	3.2	33.5	5.1

4.3 Error analysis in the high jump phase

High jump is a complex movement comprising multiple phases. In order to determine predictability of various methodologies at these phases, we split the process of high jump if into four major phases; approach run, takeoff, pole clearance, and landing. The joint angle error was calculated in every phase. Some, the results are demonstrated in box plots. The assessment of Figure 2 shows that the number of errors of all techniques is rather small and is concentrated in the period of takeoff. The observed reduction in error during the takeoff phase is likely attributable to its more standardised and repetitive biomechanical profile across athletes. Unlike the subsequent, more variable phases of flight and clearance, the takeoff action is constrained by the necessity of generating vertical impulse, resulting in a more consistent and predictable movement pattern that is

easier for the model to capture. This can be attributed to the high level of dynamic constraint of the take off movement which has relatively fixed posture patterns which are simpler to acquire. The range of error fluctuation is however much higher in all the methods during pole clearance phase. It indicates the great personal variability and creativity of the technical styles of the athletes at this stage which generates more pose uncertainty. Nevertheless, it is demonstrated that Diff-ODE has a much lower median error and dispersion than other baselines in this stage, which serves as evidence that the diffusion model component is more capable of working with and generating this diversity.

Figure 2 Distribution of joint angle errors across high jump phases, (a) acceleration phase, (b) take off phase, (c) over-the-bar phase, (d) implementation phase (see online version for colours)

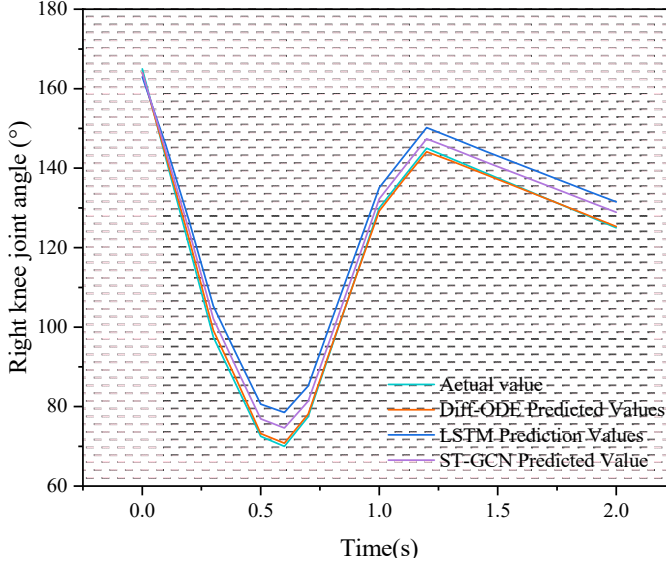


4.4 Visualisation of continuous pose prediction

To provide visual evidence of the abilities of the prediction, we chose the changes of the right knee joint angle in a test sequence and compared the actual values with diff-ODE prediction and a pair of representative baseline methods of the LSTM and ST-GCN on the same time scale. As it is evident in Figure 3, LSTM prediction curve has both visible

lag and over-smoothing. Though, the predictions made by ST-GCN come closer to the actual values the deviations are still made near the extreme points (maximum flexion angle at the moment of takeoff). On the contrary, the diff-ODE prediction curve is the most flattened and forms the nearest shape to the real values, especially in the areas of intense dynamic transitions. This justifies the benefit of neural ODEs in continuous dynamics.

Figure 3 Comparison of continuous prediction of right knee joint angle (see online version for colours)



4.5 Melting experiment

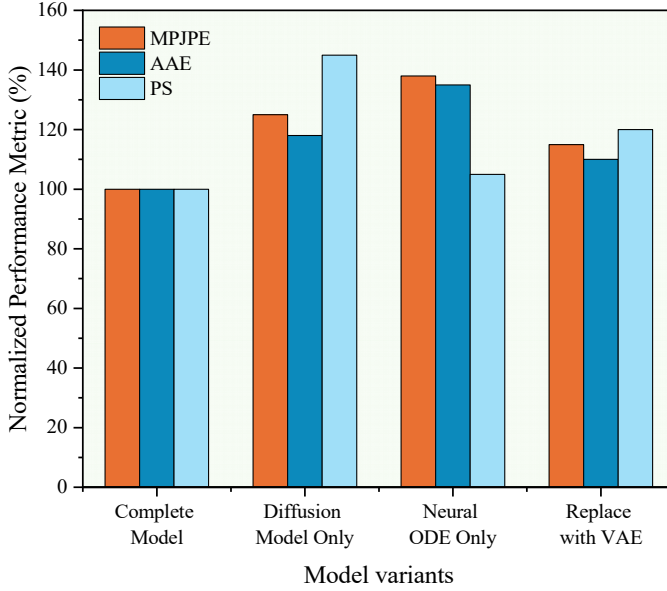
In order to explore the contributions of every component in the diff-ODE framework, we have devised ablation experiments comparing the following variant models:

- Full model: this is a model that uses both the diffusion model and neural ODE.
- Diffusion model alone: applies no further neural ODE, and further sequence smoothing is captured using linear interpolation.
- Neural ODE only: diffusion model was eliminated; raw observations directly connected to neural ODE.
- Replaced with variational autoencoder (VAE): diffusion model has been substituted by a VAE, and all other elements remain the same.

We compared these variants in three major measures (MPJPE, AAE, PS) presented in Figure 4. The complete model showed the most effective results according to all metrics, which proves the efficiency of component integration. The diffusion-only model produces realistic poses, however, with low PS, which means that there is temporal inconsistency in predicted images. The neural ODE-only variant has great smoothness but is characterised by high MPJPE and AAE, which demonstrates its inability to directly

reconstruct fine poses with noisy data because of the absence of a strong generative prior. The overall performance of performance replacement with VAE was significantly lower than that of the full version on diffusion model, which underscores the better performance of diffusion model on learning and producing intricate pose distributions. The results of these ablation experiments effectively demonstrate that the elements of the diffusion model and neural differential equation in Diff-ODE framework are mutually necessary and compliant: the diffusion model offers abundant priors of poses, whereas the neural ODE offers them to be dynamic.

Figure 4 Melting performance comparison (see online version for colours)



5 Conclusions

The paper manages to develop a complete high jump predictor architecture based on a denoising diffusion probability-based model and neural differential equations, as well as use it to validate the framework. This framework is found to be accurate in predicting the outcomes of the experiments, comprising 96.2% on the public datasets human3.6m and DIP-IMU, and makes a significant decrease of the MJPE and joint angle error to 42.6 mm and 3.2 degrees respectively. Its operation scores an overall better performance compared to various prevailing mainstream practices. What is more significant, this approach has proved to have conclusive benefits in the smoothness measures, showing that the produced posture is more compatible with real biomechanical guidelines. These numerical findings firmly support the enormous potential and special importance of integrating the generative framework with the continuous dynamical system modelling towards addressing the problem of complex motion posture prediction.

In theory, the main contribution of the current study is that a new paradigm is developed and put into practice that eliminates the conflict between the concept of

generative quality and that of dynamic continuity. The diffusion model is particularly capable of acquiring the multimodal and uncertain characteristics posed by high jump due to its mechanisms of denoising generation in steps which is much less susceptible to the averaged pose effect of traditional deterministic models. Predicting motion trajectories at discrete time scale is superseded by neural differential equations, which predict the behaviour of a dynamical system in a continuous time framework that is necessarily the description of the infinitely fine-grained process of pose change. The former works in a synergistic manner and offers rich pose priors, the latter functions as an optimiser on the temporal axis, and collaboratively provides a range of but smooth prediction results. This model provides a theoretical paradigm and methodological structure in tackling other sequential prediction issues that have the same dynamic nature like gymnastics and diving.

At the practical application level, this research delivers more precise and reliable analytical tools for sports science training and injury prevention. Coaches and athletes can leverage the system's predictions to quantitatively analyse and optimise postures during critical phases of the high jump – particularly the takeoff and pole clearance – thereby developing more targeted training regimens. Furthermore, the system's high accuracy and robustness extend its applicability to rehabilitation medicine, where it can assess athletes' recovery progress or monitor for injury-prone posture patterns. Notably, while the current model incurs relatively high computational costs, future advancements in model lightweighting and knowledge distillation techniques hold promise for deployment on portable devices or edge computing units, ultimately achieving the goal of real-time feedback.

Declarations

All authors declare that they have no conflicts of interest.

References

- Akaike, S. (2021) 'Issues and prospects on scientific advice from the viewpoint of actual policy making process', *The Journal of Science Policy and Research Management*, Vol. 36, No. 2, pp.185–192.
- Akhtar, N. and Mian, A. (2018) 'Threat of adversarial attacks on deep learning in computer vision: a survey', *IEEE Access*, Vol. 6, pp.14410–14430.
- Amat, S., Levin, D. and Yanez, R.A.F. (2024) 'A regularization-correction approach for adapting subdivision schemes to the presence of discontinuities', *Bit Numerical Mathematics*, Vol. 64, No. 1, pp.331–336.
- Aminuddin, N.A., Tajuddin, T., Mutalib, S., Dawam, S.R.M., Ali, N.R. and Din, M.M. (2024) 'Deep learning-based classification of toxic online comments using long short-term memory (LSTM) for sentiment analysis', *Journal of Information System and Technology Management*, Vol. 9, No. 37, pp.352–368.
- Amiri-Hezaveh, A., Tan, S., Deng, Q., Umulis, D., Cunniff, L., Weickenmeier, J. and Buganza Tepole, A. (2025) 'A physics-informed deep learning deformable medical image registration method based on neural odes', *International Journal of Computer Vision*, Vol. 133, No. 9, pp.6374–6399.

- Arshad, Z., Shah, Z., Raja, M.A.Z., Khan, W.A., Muhammad, T. and Ali, M. (2024) 'Novel design of artificial intelligence-based neural networks for the dynamics of magnetized chemically reactive darcy–forchheimer nanofluid flow', *Journal of Thermal Analysis and Calorimetry*, Vol. 149, No. 24, pp.15243–15276.
- Arzehgar, A., Seyedhasani, S.N., Ahmadi, F.B., Bagheri Baravati, F., Sadeghi Hesar, A., Kachooei, A.R. and Aalaci, S. (2025) 'Sensor-based technologies for motion analysis in sports injuries: a scoping review', *BMC Sports Science, Medicine and Rehabilitation*, Vol. 17, No. 1, pp.1–14.
- Bacak, A., Pnarba, A. and Dalkl, A.S. (2024) 'An experimental study on the determination of a hermetic reciprocating compressor's discharge and suction losses', *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, Vol. 238, No. 3, pp.1330–1344.
- Bintoro, K.B.Y., Priyambodo, T.K., Wicaksono, K. and Syahputra, A. (2025) 'Performance optimization of vehicle-to-vehicle communication through reactive routing protocol analysis', *International Journal of Computer Network and Information Security*, Vol. 17, No. 4, p.13.
- Caria, S. (2024) 'Labour market fragmentation and gender inequalities: a warning for deindustrializing countries, like Italy, from the case of Ecuador', *Jura Gentium*, Vol. 21, No. 2, pp.260–293.
- Deimling, C.V., Lückoff, I., Scherr, J., Catuogno, S., Mettler, N., Dreher, T. and Krautwurst, B. (2024) 'Study protocol to assess effects of fatigue on gait in children and adolescents with cerebral palsy', *Gait and Posture*, Vol. 113, No. 1, p.1.
- Dong, S., Fan, F., Chen, Y., Guo, S. and Liu, J. (2025) 'Gait planning, and motion control methods for quadruped robots: achieving high environmental adaptability: a review', *CMES-Computer Modeling in Engineering and Sciences*, Vol. 143, No. 1, pp.1–50.
- Duan, X. and Hu, Y. (2025) 'Multimodal english corpus text recognition based on unsupervised domain adaptation', *International Journal of Information and Communication Technology*, Vol. 26, No. 11, pp.53–68.
- Garcia, A.R., González, G.V. and Ruiz, R.A. (2018) 'Short-term outcome of cardiac surgery under cardiopulmonary bypass in patients who refuse transfusion: a controlled study', *The Journal of Cardiovascular Surgery*, Vol. 59, No. 5, pp.729–736.
- Gupta, A. and Kaushik, A. (2022) 'A higher-order hybrid finite difference method based on grid equidistribution for fourth-order singularly perturbed differential equations', *Journal of Applied Mathematics and Computing*, Vol. 68, No. 2, pp.1163–1191.
- Han, J.H. (2024) 'Nonequilibrium thermodynamics meets diffusion model', *Physics and High Technology*, Vol. 33, No. 12, pp.16–20.
- Huang, H., Chen, W. and Fan, Z. (2024a) 'Tscnd:temporal subsequence-based convolutional network with difference for time series forecasting', *Computers, Materials and Continua*, Vol. 78, No. 3, pp.3665–3681.
- Huang, H., Xie, L., Liu, M., Lin, J. and Shen, H. (2024b) 'An embedding model for temporal knowledge graphs with long and irregular intervals', *Knowledge-Based Systems*, Vol. 296, No. 1, p.10.
- Jiang, Y., Lu, L. and Xu, J. (2021) 'Enhanced view-independent representation method for skeleton-based human action recognition', *International Journal of Information and Communication Technology*, Vol. 12, No. 2, p.19.
- Jin, D.L., Wang, X.Y., Yuan, D.J., Li, X.D. and Du, C.Y. (2024) 'Pose prediction based on dynamic modeling and virtual prototype simulation of shield tunnelling machine', *Journal of Central South University*, Vol. 31, No. 11, pp.3854–3867.
- Li, Y., Li, S. and Qin, H.H. (2021) 'A plug-and-play scheme to adapt image saliency deep model for video data', *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 31, No. 6, pp.2315–2327.
- Mizuguchi, M., Sekine, K., Hashimoto, K., Nakao, M.T. and Oishi, S.I. (2023) 'Rigorous numerical inclusion of the blow-up time for the Fujita-type equation', *Japan Journal of Industrial and Applied Mathematics*, Vol. 40, No. 1, p.25.

- Nikolaev, E.I. (2019) 'Towards intelligent control system for computer numerical control machines', *International Workshop on Advanced Technologies in Material Science, Mechanical and Automation Engineering – MIP: Engineering: Krasnoyarsk*, Russia, Vol. 16, No. 8, pp.2–3.
- Shepherd, V., Smallman, K., Wood, F., Gillies, K., Martin, A., Moore, M., Todd, S. and Hood, K. (2025) 'Feasibility of a Study Within a Trial to evaluate a decision support intervention for families deciding about research on behalf of adults lacking capacity to consent', *Trials*, Vol. 26, No. 1, pp.1–21.
- Zhang, X.M., Han, Q.L., Ge, X. and Ding, D. (2018) 'An overview of recent developments in lyapunov-krasovskii functionals and stability criteria for recurrent neural networks with time-varying delays', *Neurocomputing*, Vol. 313, No. 3, pp.392–401.
- Zhou, X. and Tsai, C.W. (2023) 'The effects of socially shared regulation of learning on the computational thinking, motivation, and engagement in collaborative learning by teaching', *Education and Information Technologies*, Vol. 28, No. 7, p.18.
- Zhou, Y., Leng, H., Meng, S., Wu, H. and Zhang, Z. (2024) 'Structdiffusion: end-to-end intelligent shear wall structure layout generation and analysis using diffusion model', *Engineering Structures*, Vol. 309, No. 1, p.15.