Comparing MLP and LSTM Models for Human Arm Trajectory Prediction

Sofie Szlezak, Dr. Xinyao Zhang

FAMU-FSU College of Engineering

FSU Undergraduate Research Opportunity Program
Center for Undergraduate Research & Academic Engagement

FSU Florida State University

ABSTRACT

- Human trajectory prediction is important for safe and efficient human-robot collaboration.
- This study compares MLP and LSTM models for predicting arm joint motion from time-series data.
- Models were trained on the same dataset and evaluated using RMSE (root mean squared error or prediction error).
- MLP achieved substantially lower error than LSTM.
- Likely explanation: LSTMs benefit from longer motion histories than the 10-step input window used.

BACKGROUND

Data Characteristics

- Human motion prediction is commonly done using neural networks (Fig. 1).
- MLPs learn direct input-output mappings.
- LSTMs are designed to capture temporal dependencies in sequential data.
- Prior work has shown that simple MLP models can outperform sequence-based models when only short motion histories are available [2].

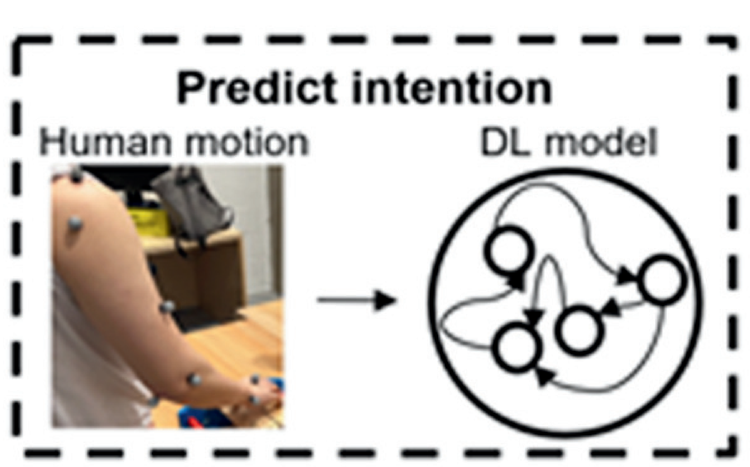


Figure 1. Conceptual illustration of human arm motion being processed by a deep learning model for movement prediction (adapted from [3]).

Research Gap

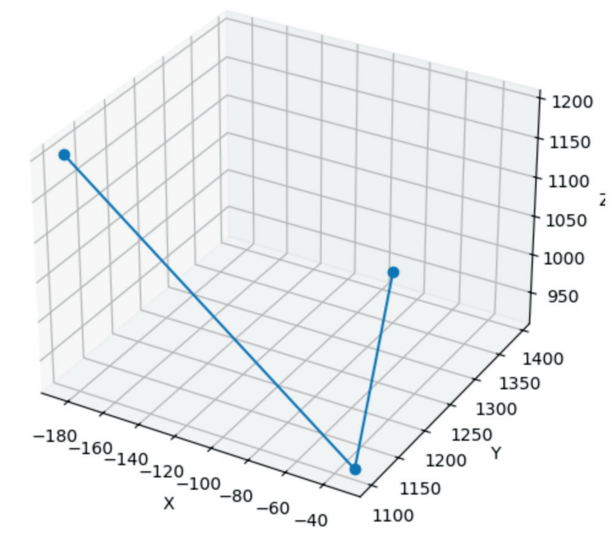
It is unclear whether sequence-based models such as LSTMs provide advantages over simpler models when only short motion histories are available.

Purpose

To compare MLP and LSTM prediction accuracy when only short motion histories are available.

METHODOLOGY

Example of one timestep/single row of motion data:



Data

- Human arm motion recordings consisting of three movement categories (A, B, C).
- Approximately 150 trials total (50 per category).

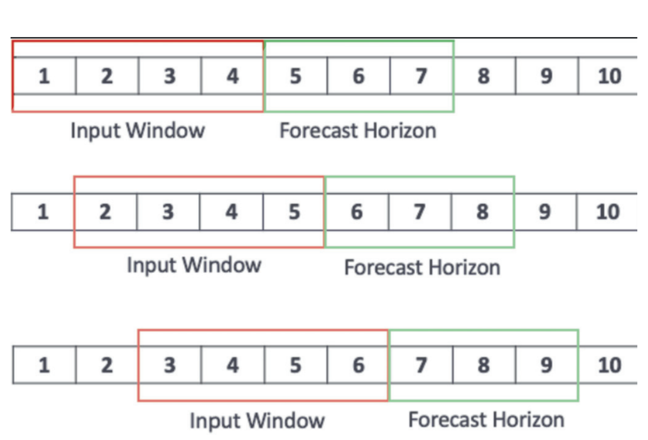
Materials / Measures

- Input: 10 consecutive timesteps of 3D joint coordinates.
- Output: Next-step joint coordinates.
- Performance measured using RMSE: average distance between predicted and actual positions

Procedure

- Sliding windows were created from each trial.
- Data were shuffled and split into training (80%) and validation (20%) sets.
- Two models were trained:
  - MLP (flattened inputs, fully connected layers)
  - LSTM (sequence-based model)

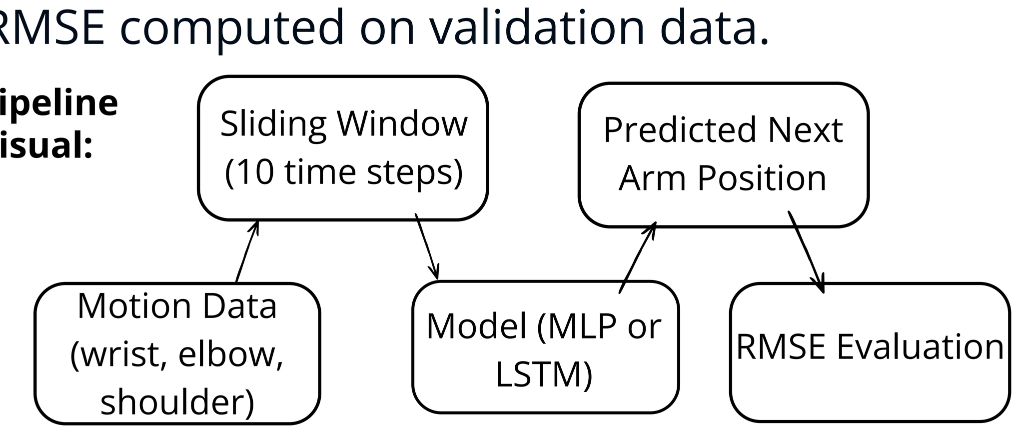
Sliding window method visual example with different numbers of input/output (adapted from [1]):



Data Analysis

- Datasets were randomly shuffled prior to training to reduce ordering bias.
- Models were trained using mean squared error loss.
- Early stopping applied to prevent overfitting.
- RMSE computed on validation data.

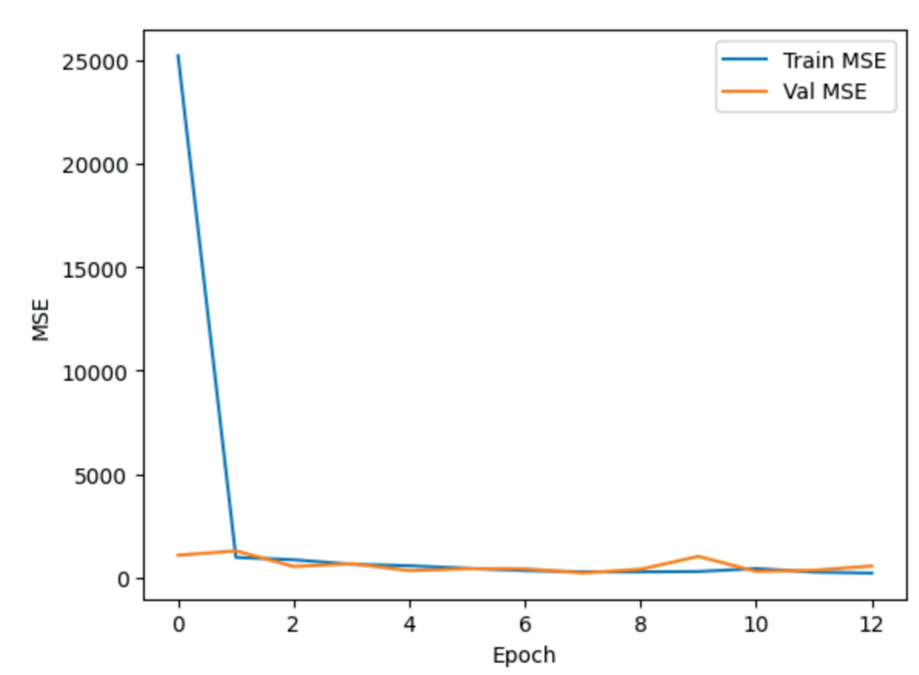
Pipeline Visual:



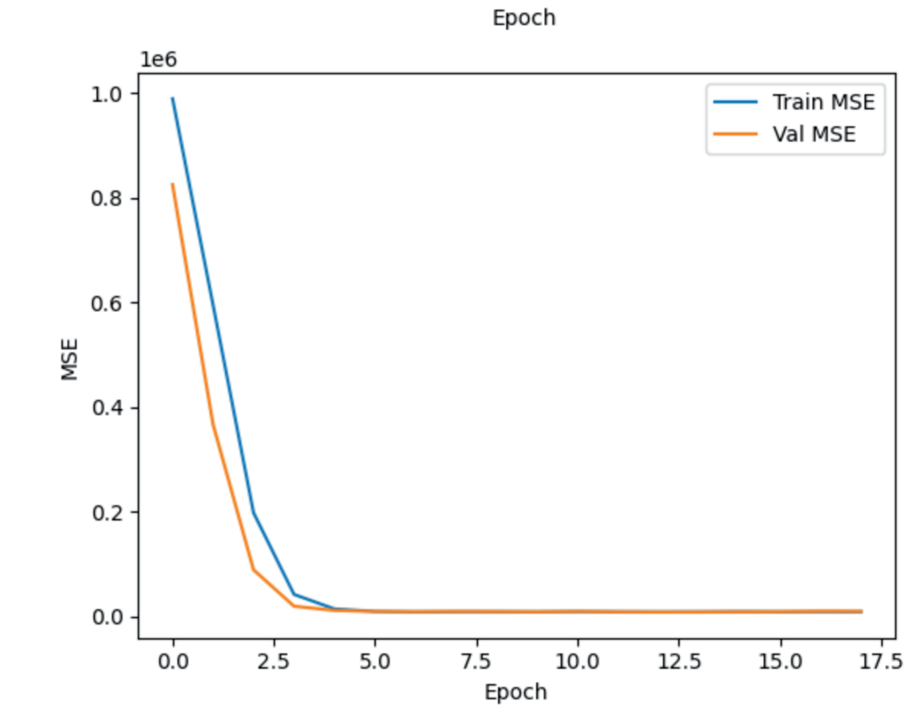
RESULTS

- Across five independent training trials, the MLP produced a mean RMSE of 12.53 ± 2.22.
- Under identical conditions, the LSTM produced a mean RMSE of 91.03 ± 4.07.
- Training and validation MSE decreased across epochs and plateaued as training progressed.
- The MLP yielded a lower RMSE than the LSTM in all five trials.

Training and Validation MSE vs. Epoch for one trial of the MLP Model:



Training and Validation MSE vs. Epoch for one trial of the LSTM Model:



DISCUSSION

The MLP outperformed the LSTM under the short input windows (10 time steps). LSTMs are built to learn patterns across longer sequences of motion, and the limited amount of past information may have reduced their advantage. With only a short motion history available, a simpler model that directly maps previous positions to the next position was sufficient enough to achieve lower prediction error.

CONCLUSIONS

- The MLP consistently produced lower prediction error than the LSTM under short input windows (10 time steps).
- The difference in RMSE between the two models suggests that model performance depends on the input history length.
- Under short input windows, simpler architectures may be preferable given the resulting lower prediction errors.

LIMITATIONS & FUTURE WORK

Limitations

- Short input windows may limit LSTM performance.
- Motion categories may not be very well balanced across training samples, potentially influencing model performance.
- The study assessed only a single input window length and a specific set of model configurations.

Future Work

- Study the impact of varying input window lengths, including longer motion histories.
- Test for effects of balancing motion categories.
- Test additional model architectures for similar trends under short input histories.
- Evaluate performance on additional motion datasets.

REFERENCES

[1] B. Benson, W. D. Pan, A. Prasad, et al., "Forecasting Solar Cycle 25 Using Deep Neural Networks," *Solar Physics*, vol. 295, no. 5, Art. no. 65, May 2020, doi: 10.1007/s11207-020-01634-y.

[2] Guo, W., Du, Y., Shen, X., Lepetit, V., Alameda-Pineda, X., and Moreno-Noguer, F., "Back to MLP: A Simple Baseline for Human Motion Prediction," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023, pp. 4798–4808.

[3] Zhang, X., Tian, S., Liang, X., Zheng, M., and Behdad, S. (January 8, 2024). "Early Prediction of Human Intention for Human–Robot Collaboration Using Transformer Network." ASME. J. Comput. *Inf. Sci. Eng.* May 2024; 24(5): 051003. https://doi.org/10.1115/1.4064258