

Pole balancing on the fingertip: Model-motivated machine learning forecasting of falls

Minakshi Debnath¹, Joshua Chang^{2, 3, *}, Keshav Bhandari¹, Dalma J. Nagy⁴, Tamas Insperger^{4,5}, John G. Milton², Anne H. H. Ngu¹,

1 Department of Computer Science, Texas State University, San Marcos, TX

2 Department of Neurology, Dell Medical School, The University of Texas at Austin, Austin, TX

3 Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, TX

4 Department of Applied Mechanics, Faculty of Mechanical Engineering, Budapest University of Technology and Economics, Budapest, Hungary

5 ELKH-BME Dynamics of Machines Research Group, Budapest, Hungary

* corresponding author

Abstract

In digital health research, there is increasing interest in developing mathematical and computational models to forecast adverse events in physiological systems. Examples include falls, the onset of fatal cardiac arrhythmias, and adverse surgical outcomes. However, the dynamics of physiological systems are known to be exceedingly complex and perhaps even chaotic. Since no model can be perfect, it becomes important to understand how the model's forecasting abilities can be improved, especially when machine learning is involved and the training data is limited. An adverse event that can be readily studied in the laboratory is the occurrence of stick falls when humans attempt to balance a stick on their fingertips. Over the last 20 years, this task has been extensively investigated experimentally and presently detailed mathematical models are available. Here we use a long short-term memory (LSTM) deep learning network to forecast stick falls. We train this model to forecast stick falls in three ways: 1) using only data generated by the mathematical model (synthetic data), 2) using only stick balancing recording of stick falls measured using high-speed motion capture measurements (human data), and 3) using transfer learning which combines a model trained using synthetic data plus a small amount of human balancing data. We observe that the LSTM model is much more successful in forecasting a fall using synthetic data than it is in forecasting falls for models trained with limited available human data. However, with transfer learning, i.e. the LSTM model pre-trained with synthetic data and re-trained with a small amount of real human balancing data, the ability to forecast impending falls in human data is vastly improved. Indeed, it becomes possible to correctly forecast 60-70 % of real human stick falls up to 2.35 s in advance. These observations support the use of physics-assisted model-generated data and transfer learning techniques to improve the ability of computational models to forecast adverse physiological events.

Introduction

There is increasing interest in leveraging physiological measurements to forecast clinically-relevant events ranging from susceptibility to cardiac arrhythmias [30, 29] to adverse surgical outcomes [4], predicting limb movements using EMG signals [26], understanding the effects of ankle weakness on human gait [24] and adverse events in human balancing tasks [14, 31, 33]. Studies over the last few years have leveraged advances in machine learning and, more recently, deep learning, to learn dynamical behaviors [7]. However, in many physiological processes, the dynamics are exceedingly complex and may not be amenable to these techniques, especially when given a limited amount of data. A case in point concerns the forecasting of falls for stick balancing on the fingertip [14]. The dynamics of the movements of the balanced stick are exceedingly complex and include on-off intermittency [2], Levy flights [3], and micro-chaotic fluctuations [19, 20]. These dynamical signatures form the cornerstone upon which current theories of human balance and falls can be based [9]. Thus it can be anticipated that strategies that forecast the fall of an inverted pendulum will translate into effective strategies for forecasting human falls. An important benchmark is how far in advance is it possible to forecast a fall? Clearly, the longer a fall can be forecasted ahead, the more time is available to activate strategies to mitigate the adverse effects of a fall [23].

Recently it was shown that by using traditional machine learning techniques (Artificial Neural Networks, Random Forest, Support Vector Machine, K-nearest Neighbour), it is possible to forecast falls for human pole balancing on the fingertip by ≈ 0.83 s [14] based on data from two pole balancers. However, it is known that machine learning model trained with small amount of data cannot generalize to the different skill levels of pole balancers, as reflected by different mean balance times. Pole balancing times can range from a few seconds to many minutes depending on the skill of the human balancer. Thus fatigue limits the number of balance trials to < 100 per day with a significant bias towards the less skilled balancers.

An attractive solution to this data scarcity problem is to use a mathematical model to generate any desired amount of synthetic stick falls and use this data to train a machine learning model to approximate the complex dynamics of stick falling. This strategy is similar in spirit to the use of physics-inspired models to investigate the effects of ankle plantar flexor weakness on gait [24], prediction of arrhythmias susceptibility[29], and to estimate limb movements using the EMG signal [26]. In all of these cases, synthetic data is generated via the relevant mathematical models in order to gain insights into the underlying physiology of the clinically-relevant system. These relevant mathematical models serve as a "digital twin" of the various systems in question[11].

We approach the difficult problem of obtaining a generalized model for forecasting of stick falls limited by a small amount of real data by first generating synthetic stick fall data using a physiologically motivated mathematical model for pole balancing on the fingertip. This mathematical model for pole balancing has been developed and refined in our laboratory over the last twenty years [Milton2023, 1, 8, 18, 19, 20, 9, 22]. Although not yet perfect, this model nonetheless reproduces the same balance time distribution and many of the frequency and kinematics-dependent properties as measured experimentally for human pole balancing. Next, we train a stacked long short-term memory (LSTM) deep learning network on synthetic data generated using this model to learn the patterns leading to stick falling. Finally, we use transfer learning to fine-tune this model with a limited amount of real-world human pole-balancing data.

LSTMs are part of a class of Recurrent Neural Network (RNN) deep-learning algorithms that have been used in time-series forecasting. Unlike traditional feed-forward neural networks, LSTMs are able to use their internal memory to process arbitrary sequences of inputs, capturing the temporal dynamics of the data. Moreover, deep learning models do not require hand-crafted feature engineering. The model can

learn the latent representation inherent in the data via lots of examples/data. Designing good hand-crafted features requires expert knowledge and is a labor-intensive process. This particular deep neural network architecture has been used previously to train a model to recognize motion-related activities from a wristwatch over a period of time so that they can be better distinguished from others [16]. A bi-directional LSTM has also been used in [5] for the forecasting of human physical activity from smartphone sensors.

We conducted an extensive set of experiments to train a range of forecasting models with forecast times ranging from 0.85 to 2.35 seconds to assess the limit of forecasting in an unstable dynamic system using a combination of measured human stick balancing data, simulated data generated by a mathematical model for stick balancing and deep transfer learning for refinement. We show that it is possible to increase the forecast time for stick falling by approximately 2-3 times as compared to the latest forecasting of impending falls by [14].

Our paper is organized as follows. First, we describe how real-world human pole balancing data is collected. Next, we present the mathematical model for human pole balancing and describe how the important parameters (time delay, sensory dead zone, and four reflex gains) were estimated. We describe how the mathematical model-generated data is pre-processed as suitable input to LSTM algorithm, and how the unbalanced nature of the fall events (i.e. falls being “rare” events relative to non-falls) was overcome. In the next section, we show how transfer learning was utilized to combine the benefits of synthetic data generation with real-world examples for improved accuracy in event anticipation. Finally, we discuss the implications of our findings and future directions for this approach for fall forecasting.

Methods

Data Collection

Human pole balancing data collection was approved by the institutional review board at Claremont McKenna College and was de-identified. Pole balancing at the fingertip was performed as described previously using the training protocol outlined in [18]. Briefly, subjects were seated in a chair while facing a blank black screen (Fig 1a). The pole is an oak dowel with diameter 6.35 mm, length 0.56 m (length of dowel plus 0.009 m reflective markers attached at each end). The mean balance time for Subject A was 31 s and for Subject B it was 22 s. Since the mean balance times were less than 240 s, both of these subjects are considered to be “novice” pole balancers [19]. The mass of the pole plus reflective markers was 13.8 g. A high-speed motion capture system (3 Qualisys Oqus 300 cameras, 250 Hz) was used to measure the position of reflective markers during pole balancing as a function of time. Balance trials in which the fall occurs in the sagittal plane were selected for this study ($\approx 70 - 80\%$ of stick falls occur in this direction [19, 9]). Data was down-sampled to 100 Hz for all our fall forecasting experiments to mimic real-world motion sensor devices, such as smartwatches and phones which typically collect data at 30-100 Hz.

Mathematical model for stick balancing

Stick balancing on the fingertip was modeled using an inverted pendulum-cart mechanical model along with human control force in the form of time-delayed predictor feedback. (Fig. 1(b)). The equation of motion for the stick itself is

$$\begin{pmatrix} \frac{1}{3}ml^2 & \frac{1}{2}ml \cos \theta \\ \frac{1}{2}ml \cos \theta & m + m_0 \end{pmatrix} \begin{pmatrix} \ddot{\theta} \\ \ddot{x} \end{pmatrix} + \begin{pmatrix} -\frac{1}{2}mgl \sin \theta \\ -\frac{1}{2}ml\theta^2 \sin \theta \end{pmatrix} = \begin{pmatrix} 0 \\ f(t) \end{pmatrix} \quad (1)$$

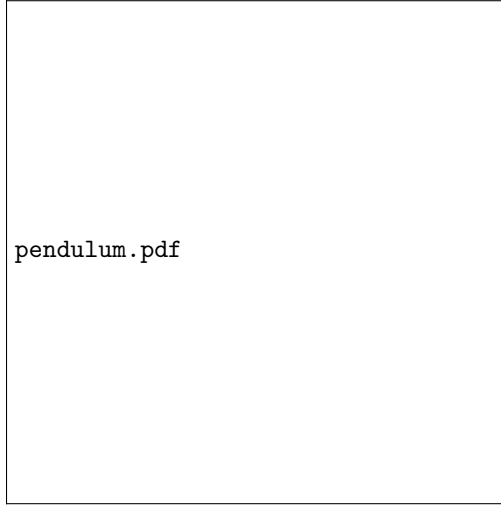


Fig 1. Human pole balancing illustrating both (a) human position during task, (b) the mathematical representation of the pole balancing behavior as an inverted pendulum-cart mechanical model.

where m is the mass of the stick, m_0 is the mass of the cart, which is equivalent to the inertia of the human arm segments plus the hand [21], θ is the vertical displacement angle of the stick, $\ddot{x}$ is the acceleration of the fingertip (cart), $\ddot{\theta}$ is the angular acceleration of the stick, and $f(t)$ is the control force. The linearized equation of fingertip motion for the control of the pendulum-cart model is

$$\begin{pmatrix} \frac{1}{3}ml^2 & \frac{1}{2}ml \\ \frac{1}{2}ml & m + m_0 \end{pmatrix} \begin{pmatrix} \ddot{\theta} \\ \ddot{x} \end{pmatrix} + \begin{pmatrix} -\frac{1}{2}mgl & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \theta \\ x \end{pmatrix} = \begin{pmatrix} 0 \\ f(t) \end{pmatrix}. \quad (2)$$

Assuming predictor feedback, the control action, $f(t)$, takes the form

$$f(t) = k_{p,\theta}\theta(t - \tau) + k_{p,x}x(t - \tau) + k_{d,\theta}\dot{\theta}(t - \tau) + k_{d,x}\dot{x}(t - \tau) + \int_{t-\tau}^t k_f(t - s)f_{PF}(s)ds, \quad (3)$$

where $k_{p,\theta}$, $k_{p,x}$, $k_{d,\theta}$ and $k_{d,x}$ are, respectively, the proportional and derivative control gains for θ and x , and k_f is a function related to the predictor feedback (for further details, see [9]). The first four terms in Eq. (3) represent the delayed state feedback, while the last term is associated with the weighted integral of the issued control force over the interval $[t - \tau, t]$ [13]. The effect of a sensory dead zone in the sagittal plane, related to difficulties with depth perception[19], was modeled as a switching component to the control. In other words, the feedback is turned on or off depending on whether θ is larger or smaller than a sensory threshold Π , i.e.,

$$\theta_{\text{perceived}}(t - \tau) = \begin{cases} 0 & \text{if } |\theta(t - \tau)| < \Pi \\ \theta(t - \tau) & \text{if } |\theta(t - \tau)| \geq \Pi \end{cases}. \quad (4)$$

However, information related to $\dot{\theta}$ and $\ddot{\theta}$ remains available [28]. The presence of this dead zone accounts for the observation that 70-80% of human stick falls occur in the anterior-posterior direction.

The model equations were integrated using semi-discretization [10] because the integral term in Eq. (3) cannot be implemented analytically. In the computer code, we

discretized the control force and approximated the integral by a discrete sum of delayed values with a sampling time of 10 ms. This approach is equivalent to the assumption that the nervous system behaves like a digital controller with a discrete input and a discrete output system. The solutions of this model exhibit micro-chaos [19, 20]

Synthetic stick falls. We consider two ways that stick falls can occur: 1) θ becomes too large and 2) x exceeds arm length. The third possibility, namely that there is a static mismatch in the gains, was not explored. For stick balancing the changes in x and θ are strongly correlated especially on time scales of $4\tau - 5\tau$. On these time scales when a change in x predicts a fall then so do changes in θ . The largest θ for which balance can be maintained depends on the skill of the stick balancer: we estimated 10 degrees for subject A and 20 degrees for subject B.

There are two constraints placed on x : 1) x cannot be larger than the effective length, L , of the subject's arm, and 2) the fingertip cannot hit the subject's chest. The effective arm length, L , is shorter than the length of the arm measured from the tip of the shoulder to the tip of the out-stretched index finger. This is a result of the "slider crank" movements of the arm which are limited by the bulk of the upper arm. These constraints were modeled by assuming that the reference position $x = 0$ of the fingertip as $L/2$. Changes in L due to the subject leaning forward were minimized by having the subject keep their back against the back of the chair at all times.

The arm length limitations become effective at time instant $t = t^*$ if either

- $x(t^*) = L/2$ and $\dot{x}(t^*) > 0$ (the arm cannot stretch further) or
- $x(t^*) = -L/2$ and $\dot{x}(t^*) < 0$ (the fingertip hits the chest).

In both cases, the fingertip position is constrained to either $x(t) = L/2$ or $x(t) = -L/2$ with $\dot{x}(t) = 0$ for $t > t^*$. This corresponds to a sudden fixation in rigid body dynamics: the pole's angular velocity changes suddenly, and the control force can be neglected compared to the arising constraint force. The pole's angular velocity after the sudden fixation (at time t^{*+}) can be calculated given that the angular momentum about the fixation point (pole's bottom) is preserved. The angular velocity just after the sudden fixation is

$$\dot{\theta}(t^{*+}) = \dot{\theta}(t^*) + \frac{2}{l}\dot{x}(t^*). \quad (5)$$

The pole's angular position after the sudden fixation can be calculated by the open-loop dynamics of a pinned pendulum. Time domain simulations show that arm length limitation ends up with the pole falling in within 1-2 seconds.

Parameter estimation. There are six parameters to be determined: the time delay (τ), the sensory threshold (Π) and the four feedback control gains ($k_{p,\theta}, k_{p,x}, k_{d,\theta}, k_{d,x}$). The time delay and threshold for the sensory dead zone were measured as described previously [19]. For all subjects we took $\tau = 0.23s$ [17, 19]. A stabilometry technique was used to estimate the control gains ([22] for more details).

We assumed that transient pole balancing resulted from alterations in the control gains for the pole angle, $k_{p,\theta}$ and $k_{d,\theta}$. The control gains for the movements for the position of the cart were not changed. Fig 2 summarizes the balance time predicted by the model as a function of $k_{p,\theta}$ and $k_{d,\theta}$ for fixed $k_{p,x}$ and $k_{d,x}$. Three types of behaviors can be observed: (1) unstable balancing with balance time ≤ 1 s, (2) stable balancing with $BT \geq 240$ s, and (3) transient balancing with $1 \text{ s} < BT < 240$ s. The blue dot shows the values of the gains for the control of the pole angle determined using stabilometry.

We used the measured mean pole balancing time to choose the gains for the control of the pole angle generated by the model (simulated data). We estimated the mean pole balancing time for simulated data by varying the initial angle from -3 degrees to 3 degrees in steps of 0.001 degrees.

BT_TX_Milton_color_v2.pdf

Fig 2. Tuning of the control gains for the simulations of the simulated data. Greyscale show the balance times, the blue dot indicates the identified control gains for balance time > 240 s, and red dot indicates the tuned control gains for balance time ≈ 31 s.

Subject	$k_{p,\theta}(N/rad)$	$k_{d,\theta}(Ns/rad)$	$k_{p,x}(N/m)$	$k_{d,x}(Ns/m)$	Π (rad)
A	120	28	14	19	1.75
B	240	30	13	30	3.00

Table 1. Control gains estimated from real pole balancing data. Note that N is Newtons, rad is radians, m is meters and s is seconds. Π refers to the dead zone threshold.

The values of the gains for the simulated data ($k_{p,x}$ and $k_{d,x}$) were chosen to give a similar mean balance time as observed experimentally for each subject (see Table 1 and Figure 3). For Subject A’s parameters, the mean balance time for simulated data was $30.4 \pm 24.8s$ (mean $\pm$ one standard deviation) and the measured mean balance time for 95 balancing trials was $31 \pm 19.8s$. The red dot in Fig. 2 shows the values of the gains. For Subject B’s the mean balance time for simulated data was $24.6 \pm 9.8s$ and the measured mean balance time for 21 balancing trials was $24.8 \pm 13.05s$.

Data generation and pre-processing for machine learning. We configured the mathematical model using the control gains shown in Table 1. An initial angle was randomly selected from a range between $[-3,3]$. The data generation process involves running six simulations, and computing the feature set: fingertip position x , and vertical displacement angle θ at each time step. A total of 1000 balancing trials are generated in each simulation. The time point at which balance control was lost (LCT) was determined as follows: First, the largest θ for which the stick remained balanced was determined from all of the balance trials. Balance control was lost when the angle in the next time step (0.01s) was larger. The data points for times greater than LCT were discarded. This is indicated to the right of the vertical orange line in Fig. 3 Since our focus is on forecasting, not fall detection, we further removed the 0.75s prior to LCT to make sure that the changes in angular displacement θ and fingertip position x were not those related to falling (i.e. truly pre-fall data). This time can also be thought of as time (buffer zone) available to activate strategies to either prevent or minimize the consequences of a fall.

The remaining data is then partitioned into no-fall and pre-fall regions and further segmented into fixed-size windows creating training instances/windows using the sliding

window protocol as shown in Fig. 3. In all of our experiments, each window is set to 1.28 s long which is sufficient to account for the dynamics of a pre-fall pattern. The amount of data labeled as pre-fall is determined by the time-to-predict parameter and the size of the window we used. We vary the time-to-predict parameters from 0.1 to 1.5s to assess how far ahead we can forecast impending fall in an unstable dynamical system.

images/figure_3_up.pdf

Fig 3. Comparison of the changes for the last 10s of balancing in vertical displacement angle, θ , for synthetic data and real pole balancing data for Subject A. The data pre-processing process excluded the data points in the last portion of the signal (labeled as loss of control, LCT region) that correspond to the start of a fall. We divide the remaining data into no fall and pre-fall which we label as 0 and 1. The generated data was further partitioned into windows, which are visually represented by boxes colored in shades of gray (indicates no-fall region) and brown (indicates pre-fall region). The horizontal dashed line indicates the θ when the control was lost.

Balancing training data. Given that each simulation will generate a lot more data for the no-fall region than the pre-fall region, our data set is massively unbalanced. In machine learning, training with an unbalanced dataset will end up with a model that cannot be generalized. To balance our data set, we dynamically computed the stride size within the no-fall region to match the number of windows within the pre-fall region during the sliding window generation process. We present this innovative balancing

process in Fig 4. The key is to first maintain a fixed stride size of 1 (or other suitable sizes) within the pre-fall region. Assuming that the length of the pre-fall region is represented by n , we will obtain a total of $(n - w + 1)$ sliding windows within the pre-fall region, with a stride of 1 and a window size of w . Afterward, we dynamically calculate the stride size ($S1$) within the no-fall region to correspond to the number of windows in the pre-fall region using this computation:

$$S1 = \frac{(\text{number of sliding windows in no fall region})}{(\text{number of sliding windows in pre-fall region})}$$

where, the number of sliding windows in the no-fall region is $(N - w + 1)$. Here N refers to the length of the no-fall region. By adjusting the stride size within the no-fall region, the model observes a nearly equal number of data windows from both the no-fall and pre-fall regions.

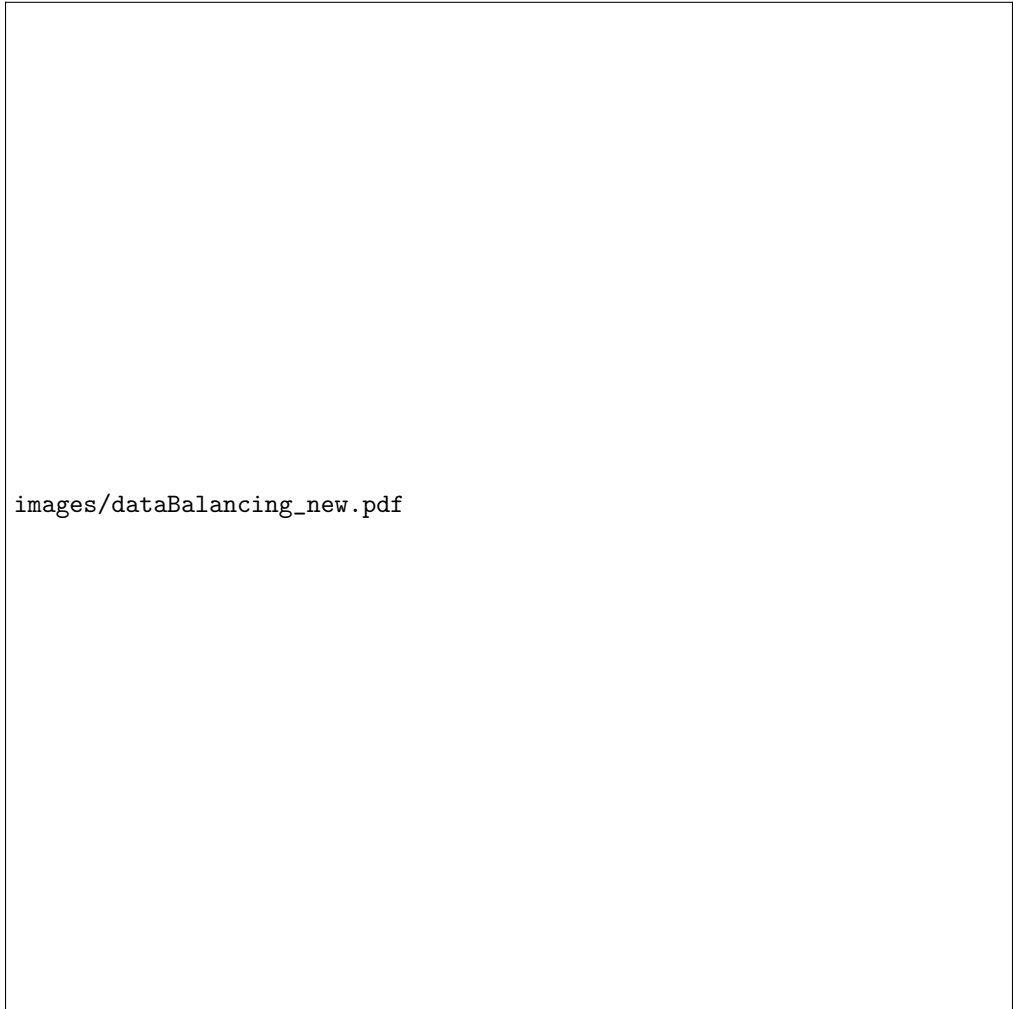
The traditional or naive way of balancing the dataset is by running additional simulations, labeling the simulated data as no-fall and pre-fall after cropping off the specified number of data points in the LCT region at the end, and only saving the pre-fall section of the data for training. This is a much more computation-intensive way of data balancing and might generate inconsistent training datasets since additional pre-fall data do not have a corresponding no-fall data region. Given the temporal relationships that could arise between the corresponding no-fall and pre-fall regions, only adding in one of those regions could negatively affect the training process.

LSTM Networks and Model Training

The ideal model for fall forecasting should be able to learn a nonlinear function that could accurately recognize patterns preceding falls, within a specified amount of time ahead (i.e. time to predict). In all our experiments, we used a stacked bidirectional LSTM deep learning model which can capture complex dependencies between the sequential nature of pole balancing data. The distinct structure of LSTM facilitates the selective preservation of crucial past information while ignoring nonessential information. Our choice is supported by other researchers. For example, in [31], many variants of stacked LSTM models were used for crash prediction in a disorienting spaceflight analog balancing task from the time series data. The utilization of the bi-directional LSTM model can result in a notable increase in the quantity of information accessible to the network, consequently enhancing the contextual cues provided to the algorithm. Fig-5 shows the stacked bi-directional LSTM architecture we used with batch normalization. Incorporating a batch normalization layer stabilizes the training process, facilitating faster convergence, and enhancing model performance. Moreover, it also regularizes the model, minimizing overfitting and enhancing its ability to generalize to new data to make the model more robust.

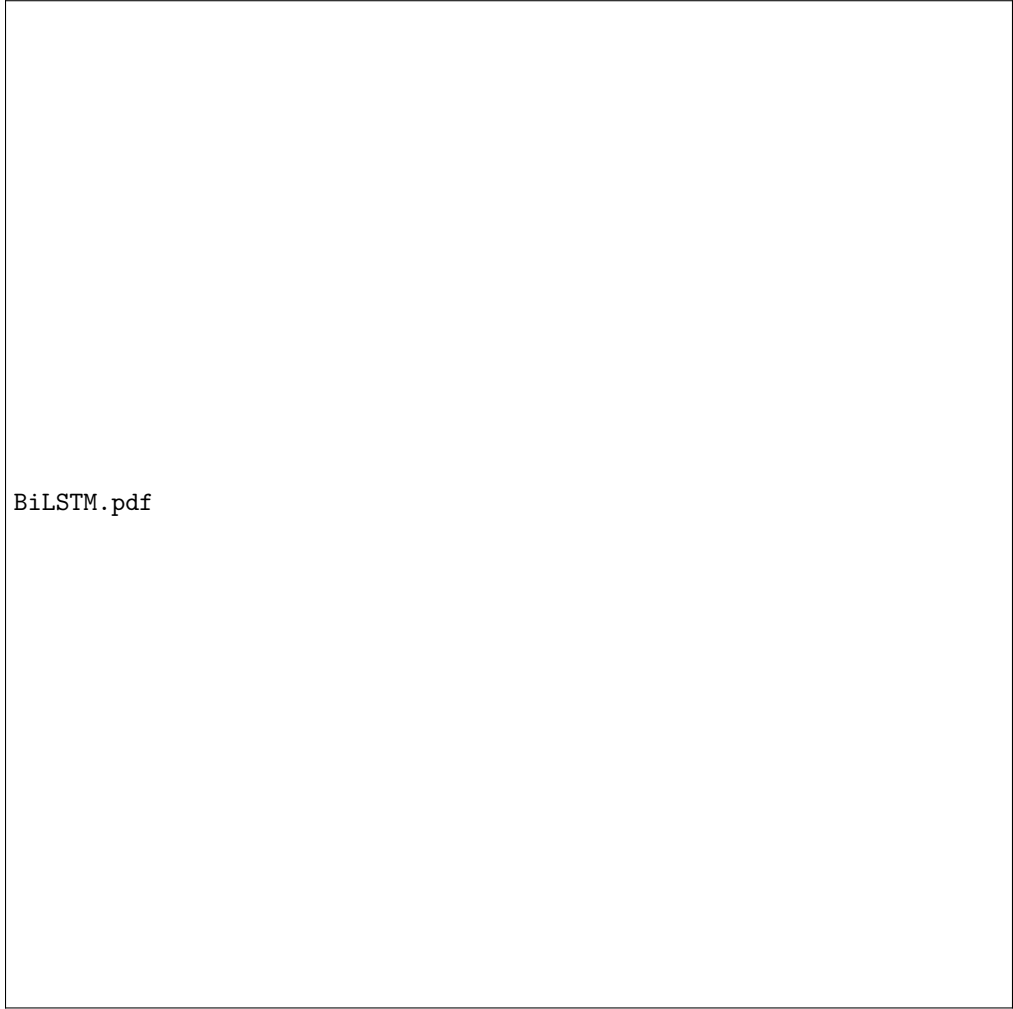
We selected SGD (Stochastic Gradient Descent) as our optimizer since it can lead to faster convergence, which is vital for quickly learning from historical data in time series forecasting. Since our problem can be framed as a pre-fall-vs-no-fall classification, a Binary-Cross-Entropy (BCE) is chosen for the loss function. We set a large epoch size of 100 but adopt an early stopping strategy. All our models were built and trained using Keras and TensorFlow on Dell Precision 7820 Tower with 256 GB RAM, and four GeForce GTX 1080 GPUs. The parameters used in our bi-directional LSTMs are presented in Table 2.

Input Features. The inputs to our bi-directional LSTM are the angular displacement and the position of the pole both at time t and at $t - \tau$, where τ is the time delay in the mathematical model as discussed earlier. The vertical angular displacement is defined as the angle made by the pole with the sagittal plane. Similarly,



images/dataBalancing_new.pdf

Fig 4. Computing stride size for no fall region dynamically to balance the data



BiLSTM.pdf

Fig 5. A bidirectional recurrent unit (LSTM) for fall forecasting.

the position of the pole at time t is defined by the actual position of the fingertip (on the x-axis). Data is fed one window at a time with the above four features. It is important to note that in the experiments, the dimension of the input features is defined as, $X_{input} = (\theta(t - \tau), \theta(t), x(t - \tau), x(t))$ such that $X_{input} \in \mathbb{R}^{d \times 4}$, where d is the size of the window of data (128 or 1.28 s) for the optimal learning of pre-fall signals.

Model Validation. To determine how well our models perform and which model will perform best on unseen data, we created a training and validation data set with an approximate split of 90% of data for training and the other 10% for testing. The data is split in such a way that none of the data in the validation set appears in the training set. Data used for our experiments are balanced in terms of no-fall and pre-fall data windows. Since time to predict will affect the amount of data available for training, down-sampling of data is performed to ensure that all our experiments are using the same number of windows of no-fall and pre-fall data as used in the experiment using the smallest time-to-predict value.

Evaluation Metric. We adopted the standard metric to evaluate the performance of a machine-learning model. There are the F1 score, Precision, Recall, and Accuracy.

Table 2. LSTM Model Configuration

Names	BiLSTM
Learning Rate	0.001
Epochs	100
Batch Size	128
Optimizer	SGD
Loss Function	BCE

These metrics are defined as:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1_Score = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

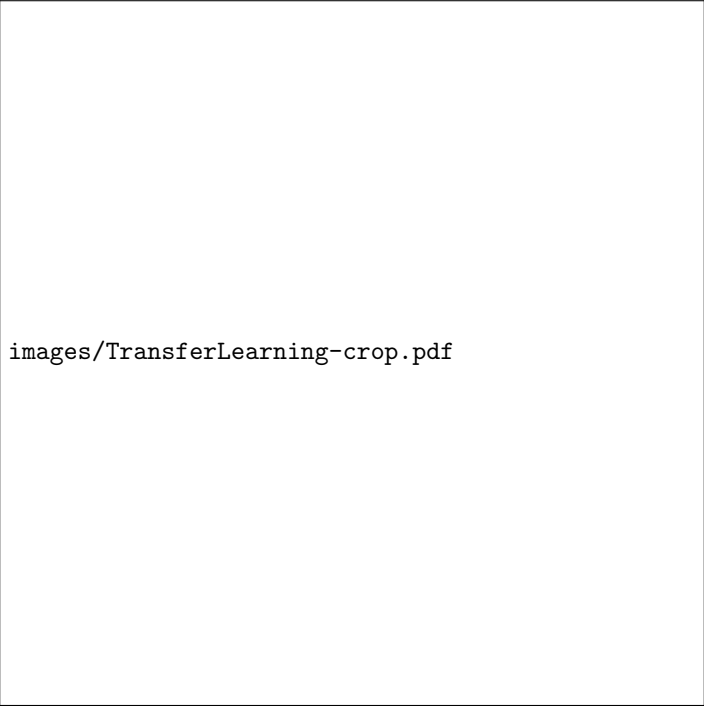
True Positive(TP) occurs when the model correctly predicts a positive instance, such as accurately forecasting a fall. True Negative (TN) is when the model correctly predicts a negative instance (non-fall for our case). False Positive is the case when the model erroneously predicts a negative data sample as positive and False Negative is the opposite case. Precision, which is the same as sensitivity, assesses the accuracy of positive prediction, and Recall, which is the same as specificity, quantifies the correct identification of real positives. The F1 Score is a harmonic mean of precision and recall to provide a balanced measure of model performance.

Transfer Learning. In the computer vision world, transfer learning has been applied successfully to transfer knowledge from a model trained in an existing domain to data captured from a new domain that has a slightly different prediction target. For example, using transfer learning, you can take an existing algorithm trained to identify 50 different flowers in images and use it to create a new algorithm that can identify those 50 flowers plus 10 new types of flowers, all while reducing the training time of the new algorithm [StanfordTL-2021, 25]. Essentially, transfer learning uses a pre-trained model as a base for a new model of a target in a similar domain. Transfer learning allows us to refine a base model with a small amount of real human pole balancing data set and then evaluate the model on the rest of the real data. Fig 6 is a schematic diagram that illustrates this process.

We implemented the transfer learning by first initializing the network with the pre-trained weights of a saved model. We split our real pole balancing data where one set was used for the refinement of the saved model and another set was used for testing the model after refinement.

Experiments

LSTM training with only real data. At first, we trained our LSTM model using only measured data captured by the motion camera from the human subject. This is to create a baseline model for comparison with other models in our experiments. We used two subjects' data to conduct our experimental study. For subject A, there were a total of 28 data samples available. We used 26 data samples for training and 2 samples for testing. For subject B, only 14 samples were available. We thus used 12 data samples



images/TransferLearning-crop.pdf

Fig 6. Comparison of traditional Machine Learning versus Transfer Learning

for training and 2 for testing. We only trained one forecasting model using the smallest time to predict value (0.1 s) for each subject. Table 3 shows the result of this experiment for both subjects A and B's real pole balancing data.

LSTM training with only synthetic data. The goal of this experiment is to train a series of models purely using synthetic data with different time-to-predict values to assess the limit of the performance of the LSTM machine learning algorithm for predicting impending stick falls. A MATLAB program that implemented the mathematical model described in [section](#) was used to generate synthetic data for subjects A and B using the estimated parameter shown in Table 1. The detailed implementation of the mathematical model is described in the supplementary section. We labeled the generated data set into no-fall and pre-fall regions, segmented them into windows of 1.28 s long, and balanced the no-fall and pre-fall windows as described earlier in [section ??](#). In each of the LSTM training sessions, the training set contained 1000 samples, while the validation set contained 50 samples. This model is then tested on the small amount of real pole balancing available to us to measure how the model trained with synthetic data performs against real data. For each subject, we repeated this LSTM training six times with different batches of synthetic training data to eliminate the probability of the randomness effect. In addition, each of these LSTM training is run 15 times using a different time-to-predict value from 0.1 s to 1.5 s with a step size of 0.1.

To ensure that our results are not due to insufficient training data, we trained one LSTM model using 10,000 balance trials of synthetic data. Training this large data set took eight hours on a computer with four GPUs. The model trained using this large data set performed similarly to the one trained with only 1000 balance trials. For computational efficiency, we ran all of our remaining experiments with only 1000 balance trials.

Refining LSTM model via transfer learning. In this third experiment, we

studied the impact of transfer learning on improving a model trained using purely synthetic data by utilizing a small amount of real human pole-balancing data. To achieve that, we refined all the models trained with synthetic data using both subject A and B’s real-human pole balancing data. As in other experiments, we split the available human balancing data into training and validation sets. The model was re-trained with a very small learning rate so that we do not drastically change the trained weights.

Results

Training with only real data.

Table 3 presents the result of training solely on real human pole balancing data, which reveals bad performance for both subjects A and B. This data was captured using only a time-to-predict of 0.1 s, which presumably should be the best that can be done as the data is closest to the fall. These results were expected as the limited availability of data was not able to train the LSTM model effectively.

	Subject A	Subject B
Precision	0.48	0.38
Recall	0.37	0.40
Accuracy	0.55	0.54

Table 3. Result of training with only a small amount of real data. Time-to-predict is 0.1 s.

Training with only simulated data.

In figure 7, we present the average and standard deviation of the precision, recall, and accuracy values obtained from six independently trained LSTM models for both subject A’s and B’s simulated data. The LSTM models performed well when tested using simulated data, but did not perform as well when tested against real human data. For simulated data, we observe that as the time-to-predict values increase, there is a decrease in accuracy, precision, and recall scores. This trend holds true for accuracy and precision scores when dealing with real data as well. This observation applies to both Subject A’s and Subject B’s data. These results indicate that our LSTM model can perform better with a shorter time to predict value, which is expected given that the larger time-to-predict values incorporate data further away from the fall itself. The recall score for both Subject A and B’s real data appears to exhibit some variability with respect to the time to predict values. One possible explanation for this low recall could be that the simulated data may not capture the full range of diversity and complexity found in real pole-balancing scenarios. Additionally, the model trained on simulated data may not exhibit effective generalization when applied to real pole-balancing data.

Refining LSTM model through transfer learning

Figure 8 summarizes the results of the LSTM model trained initially on simulated data and refined with real human pole balancing data. As we can see from this figure, the results for both subjects A and B suggest an improvement with transfer learning. We see that for both subjects, precision, recall, and accuracy have all improved substantially from when the LSTM model was trained only on simulated data. After

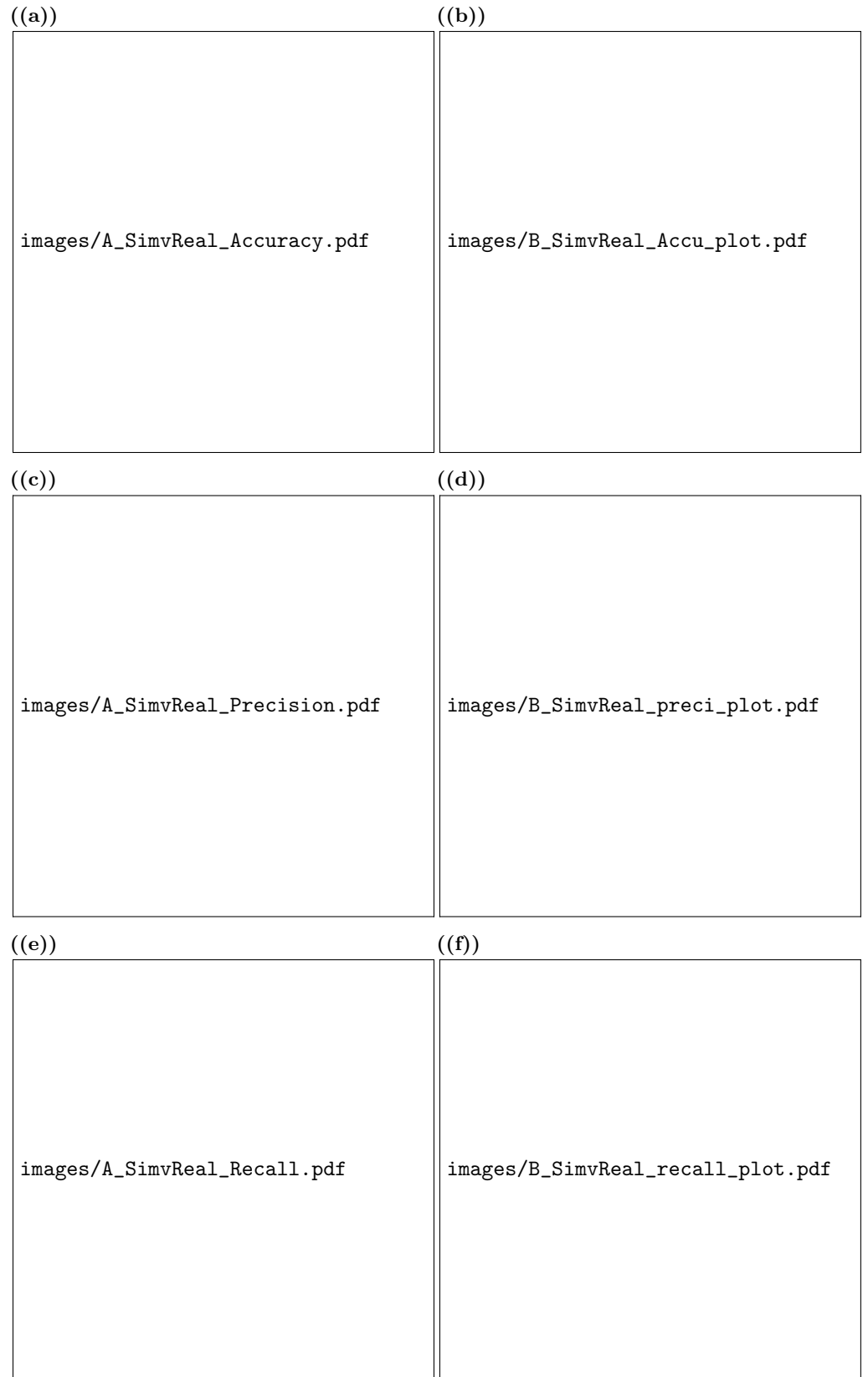
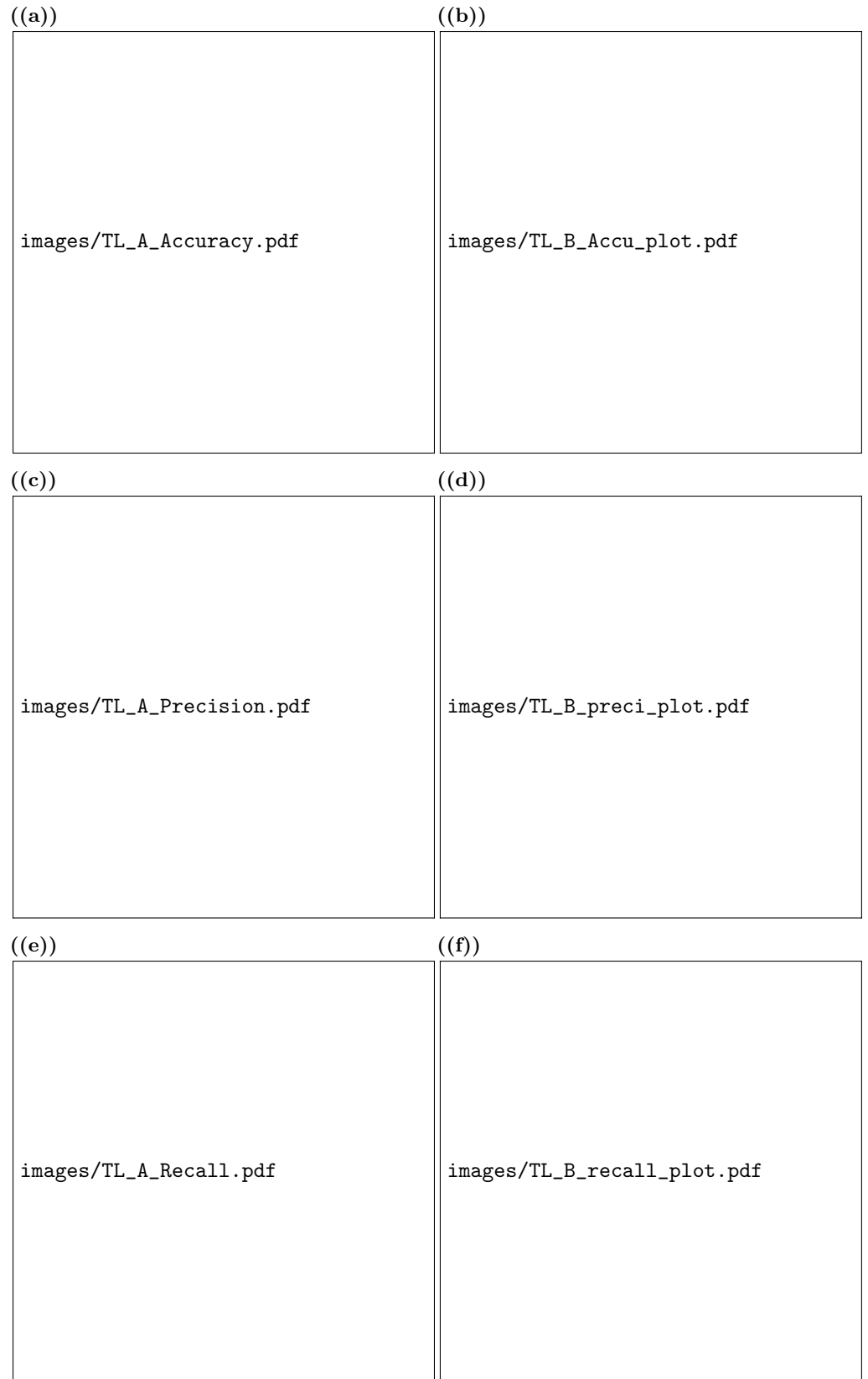


Fig 7. Results for training with only simulated data. In each plot, test results on simulated data (blue) are compared to results when tested with real human data (orange). Both data from subject A (a-c) and subject B (d-f) are displayed. Each data point represents the mean accuracy (a, d), precision (b, e), and recall (c, f) values evaluated across six separate trained models. The bars indicate the standard deviation.



implementing transfer learning, the overall accuracy for both Subject A and B exhibited an improvement of approximately 10% to 18%. The results also demonstrate an increase in both precision and recall for both subjects. These improvements indicate that the LSTM model, which was originally trained solely on simulated data, now demonstrates improved generalization to real pole balancing data.

When evaluating accuracy and precision scores on data from both subjects, the results after implementing transfer learning also reveal a subtle pattern where the scores gradually decrease as the time to predict values increases. This phenomenon is expected as an increasing time-to-predict means that the window being examined is farther away from the fall. As with any forecasting problem, there is more uncertainty when predicting far into the future compared to predicting closer to the present. This is even more true with complex, chaotic dynamical systems as the smallest fluctuations can have rapid divergence in a few time steps.

We notice that while our accuracy and precision are relatively high, our recall is low when testing against real human balancing data. In our classification, a threshold of 0.5 is being used. Recall can be improved by fine-tuning the hyper-parameters of the LSTM model such as the window size, and the systematic investigation of the trade-off between precision and recall in future work.

Discussion

While traditional machine learning techniques have been used by many researchers for the forecasting of adverse events [29, 14, 31, 33], recent studies have shown that when data sets are complex and dynamic, deep machine learning techniques always surpass these approaches [15]. In our previous study, we compared traditional machine learning techniques (Support Vector Machine, Random Forest, Naive Bayes) against deep learning architectures for fall detection and found better performance from deep learning approaches [16]. In this paper, we investigated in depth the use of transfer learning and a stacked LSTM deep learning model to train the forecasting model for stick fall.

We addressed the data scarcity problem in trying to forecast impending falls using deep learning by showing that data generated by a mathematical model for stick balancing can be used to effectively train a deep machine learning algorithm to predict stick falls. We balance the massively unbalanced data using an innovative balancing technique. In particular, our results show that by combining mathematically generated data with a limited amount of real data using transfer learning, the ability to forecast falls is increased significantly over that obtained with training using a small amount of available real data alone. Our approach resulted in a fall forecasting model that is 60-70% reliable up to 2.35 seconds. This increase in forecasting times increases the time available for activating strategies to minimize the effects of a fall [14, 23].

As stated earlier, we noticed that as the time-to-predict increases, the accuracy and prediction decrease. One way to understand this is that the latent space representations of the pre-fall and no-fall data windows become more similar to some extent and that the "signatures" indicative of the dynamics of pre-fall may not have fully manifested because the time-to-predict is too large. This provides us with an interesting direction for future research in exploring the specific dynamical changes that occur as the fall approaches. One possible future direction is thus to train our model with a contrastive supervised learning[12] approach. By this method, we first train our model to distinguish between the pre-fall data and no-fall data by increasing the distance between the latent space representation and later refine the model by further training to classify the pre-fall and no-fall. Another direction is to compare the effectiveness of the forecasting model trained with the mathematically generated data versus other techniques such as Diffusion [27], GAN [6], and statistical approaches when data

scarcity is a problem. This will confirm the importance of empirical modeling to capture "key signatures" of stick fall as compared to those reproduced by the popular generative AI approaches.

The results that we have when tested against simulated data indicate that the stacked LSTM architecture is able to accurately forecast complex and chaotic dynamical systems, however, when we tested against real human data, the low recall suggests that our model can continue to be improved in its anticipation of a fall. However, the precision indicates that when the model suggests a fall is indeed impending, we can expect a fall to occur. Overall, for about 30% of the time, the models is however unable to distinguish between pre-fall and no-fall windows. These results suggest that this approach has indeed improved our ability to anticipate fall events despite the complexity and chaotic nature of the data, but that discrepancies between the simulated data and real data exist.

Indeed, in our approach, the matching of the mathematical model to human data was purposely designed to take into account the individual's differences in mean balance time, the gains for time-delayed predictive feedback, and their sensory dead zones. However, when plotting the real human data and synthetic data side by side, we can visualize the differences in their dynamical properties as shown in Figure 3. This suggests that there may be other factors important to anticipating impending stick falls that the LSTM has learned to anticipate, that are not necessarily found in the real data.

The above points are critically important since many attempts to generate synthetic data could potentially fail for the same observation. For example, although synthetic data generation by Generative Adversarial Networks (GANs) could potentially construct time-series data that have similar statistical properties as real data when aggregated, GANs may fail to accurately capture key features or signatures at a single trajectory level that are indicative of an impending collapse or critical transition. Without having some insights into the appropriate metrics that describe these key features, synthetic data generators may instead construct data that are very accurate in every feature except the signatures themselves. The fact that human pole balancers are able to sense when they are about to lose control suggests that a signature does exist, and thus that there are features that should be matched between synthetic data and real human data. Future directions may explore the use of alternative metrics to match real data with those generated through mathematical models including decomposition-based, statistical generative models and alternative embedding spaces[32].

Acknowledgements

We thank Becca Shope for assistance in collecting the human stick balancing data. The research in this paper is supported by the National Science Foundation under the NSF-SCH grant (2123749). It is also supported by the National Research, Development and Innovation Fund (Grants BME-NVA-02 and TKP2021-EGA-02) provided by the Ministry of Innovation and Technology financed under the TKP2021 funding scheme and by the National Research, Development and Innovation Office (Grant no. NKFI-K138621). JM acknowledges [past](#) support from the William R. Kenan Jr Charitable Trust.

References

- [1] J. L. Cabrera and J. G. Milton. "Stick balancing: On-off intermittency and survival times". In: *Nonlinear Studies* 11 (2004), pp. 305–317.

- [2] Juan L. Cabrera and John G. Milton. “On-Off Intermittency in a Human Balancing Task”. In: *Physical Review Letters* 89.15 (Sept. 2002), p. 158702. DOI: 10.1103/PhysRevLett.89.158702.
- [3] Juan Luis Cabrera and John G. Milton. “Human Stick Balancing: Tuning Lévy Flights to Improve Balance Control”. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 14.3 (Sept. 2004), pp. 691–698. ISSN: 1054-1500. DOI: 10.1063/1.1785453.
- [4] Hugh Chen et al. “Forecasting Adverse Surgical Events Using Self-Supervised Transfer Learning for Physiological Signals”. In: *npj Digital Medicine* 4.1 (Dec. 2021), pp. 1–13. ISSN: 2398-6352. DOI: 10.1038/s41746-021-00536-y.
- [5] Emre Çiçek, Sezer Gören, and Gokhan Memik. “Physical Activity Forecasting with Time Series Data Using Android Smartphone”. In: *Pervasive and Mobile Computing* 82 (June 2022), p. 101567. ISSN: 1574-1192. DOI: 10.1016/j.pmcj.2022.101567.
- [6] Yaroslav Ganin et al. “Domain-Adversarial Training of Neural Networks”. In: *Journal of Machine Learning Research* 17.59 (2016), pp. 1–35. DOI: 10.1007/978-3-319-58347-1_10.
- [7] William Gilpin, Yitong Huang, and Daniel B. Forger. “Learning Dynamics from Large Biological Data Sets: Machine Learning Meets Systems Biology”. In: *Current Opinion in Systems Biology* 22 (Aug. 2020), pp. 1–7. ISSN: 24523100. DOI: 10.1016/j.coisb.2020.07.009.
- [8] Tamas Insperger and John Milton. “Sensory Uncertainty and Stick Balancing at the Fingertip”. In: *Biological Cybernetics* 108.1 (Feb. 2014), pp. 85–101.
- [9] Tamás Insperger and John Milton. *Delay and Uncertainty in Human Balancing Tasks*. Lecture Notes on Mathematical Modelling in the Life Sciences. New York, NY: Springer International Publishing, 2021. DOI: 10.1007/978-3-030-84582-7.
- [10] Tamás Insperger and Gábor Stépán. *Semi-Discretization for Time-Delay Systems: Stability and Engineering Applications*. Springer Science & Business Media, July 2011. ISBN: 978-1-4614-0335-7.
- [11] Michael G Kapteyn, David J Knezevic, and Karen Willcox. “Toward predictive digital twins via component-based reduced-order models and interpretable machine learning”. In: *AIAA Scitech 2020 Forum*. Orlando, FL: American Institute of Aeronautics and Astronautics, Jan. 2020. ISBN: 978-1-62410-595-1. DOI: 10.2514/6.2020-0418. (Visited on 10/06/2023).
- [12] Prannay Khosla et al. “Supervised Contrastive Learning”. In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 18661–18673.
- [13] Miroslav Krstic. *Delay Compensation for Nonlinear, Adaptive, and PDE Systems*. Systems & Control: Foundations & Applications. Boston: Birkhäuser Boston, 2009. DOI: 10.1007/978-0-8176-4877-0.
- [14] Miguel Rodríguez Márquez et al. “Machine Learning Forecasting of Extreme Fluctuations in a Human Balancing Task”. In: *Knowledge-Based Systems* 280 (Nov. 2023), p. 111000. ISSN: 0950-7051. DOI: 10.1016/j.knosys.2023.111000.
- [15] Gerard Martí-Juan, Gerard Sanroma-Guell, and Gemma Piella. “A Survey on Machine and Statistical Learning for Longitudinal Analysis of Neuroimaging Data in Alzheimer’s Disease”. In: *Computer Methods and Programs in Biomedicine* 189 (June 2020), p. 105348. ISSN: 0169-2607. DOI: 10.1016/j.cmpb.2020.105348.

- [16] Taylor R. Mauldin et al. “SmartFall: A Smartwatch-Based Fall Detection System Using Deep Learning”. In: *Sensors* 18.10 (Oct. 2018), p. 3363. ISSN: 1424-8220. DOI: 10.3390/s18103363.
- [17] Biren Mehta and Stefan Schaal. “Forward Models in Visuomotor Control”. In: *Journal of Neurophysiology* 88.2 (Aug. 2002), pp. 942–953. ISSN: 0022-3077. DOI: 10.1152/jn.2002.88.2.942.
- [18] John Milton and Tamas Insperger. “Acting Together, Destabilizing Influences Can Stabilize Human Balance”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 377.2153 (July 2019), p. 20180126. DOI: 10.1098/rsta.2018.0126.
- [19] John Milton et al. “Control at Stability’s Edge Minimizes Energetic Costs: Expert Stick Balancing”. In: *Journal of The Royal Society Interface* 13.119 (June 2016), p. 20160212. DOI: 10.1098/rsif.2016.0212.
- [20] John G. Milton et al. “Microchaos in Human Postural Balance: Sensory Dead Zones and Sampled Time-Delayed Feedback”. In: *Physical Review E* 98.2 (Aug. 2018), p. 022223. DOI: 10.1103/PhysRevE.98.022223.
- [21] Dalma J. Nagy, László Bencsik, and Tamás Insperger. “Experimental Estimation of Tactile Reaction Delay during Stick Balancing Using Cepstral Analysis”. In: *Mechanical Systems and Signal Processing* 138 (Apr. 2020), p. 106554. ISSN: 0888-3270. DOI: 10.1016/j.ymssp.2019.106554.
- [22] Dalma J. Nagy, John G. Milton, and Tamas Insperger. “Controlling Stick Balancing on a Linear Track: Delayed State Feedback or Delay-Compensating Predictor Feedback?” In: *Biological Cybernetics* 117.1 (Apr. 2023), pp. 113–127. ISSN: 1432-0770. DOI: 10.1007/s00422-023-00957-w.
- [23] Banne Nemeth et al. “Prevention of Hip Fractures in Older Adults Residing in Long-Term Care Facilities with a Hip Airbag: A Retrospective Pilot Study”. In: *BMC Geriatrics* 22.1 (June 2022), p. 547. ISSN: 1471-2318. DOI: 10.1186/s12877-022-03221-1.
- [24] Carmichael F. Ong et al. “Predicting Gait Adaptations Due to Ankle Plantarflexor Muscle Weakness and Contracture Using Physics-Based Musculoskeletal Simulations”. In: *PLOS Computational Biology* 15.10 (Oct. 2019), e1006993. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1006993.
- [25] Sinno Jialin Pan and Qiang Yang. “A Survey on Transfer Learning”. In: *IEEE Transactions on Knowledge and Data Engineering* 22.10 (Oct. 2010), pp. 1345–1359. ISSN: 1041-4347. DOI: 10.1109/TKDE.2009.191.
- [26] Yaroslav Smirnov et al. “Solving Musculoskeletal Biomechanics with Machine Learning”. In: *PeerJ Computer Science* 7 (Aug. 2021), e663. ISSN: 2376-5992. DOI: 10.7717/peerj-cs.663.
- [27] Yusuke Tashiro et al. “CSDI: Conditional Score-based Diffusion Models for Probabilistic Time Series Imputation”. In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 24804–24816.
- [28] Andreas Thiel et al. “Contribution of Individual Retinal Ganglion Cell Responses to Velocity and Acceleration Encoding”. In: *Journal of Neurophysiology* 98.4 (Oct. 2007), pp. 2285–2296. ISSN: 0022-3077. DOI: 10.1152/jn.01342.2006.
- [29] Meera Varshneya, Xueyan Mei, and Eric A. Sobie. “Prediction of Arrhythmia Susceptibility through Mathematical Modeling and Machine Learning”. In: *Proceedings of the National Academy of Sciences* 118.37 (Sept. 2021), e2104019118. DOI: 10.1073/pnas.2104019118.

- [30] Mark A. Walker et al. “Estimating the Probabilities of Rare Arrhythmic Events in Multiscale Computational Models of Cardiac Cells and Tissue”. In: *PLOS Computational Biology* 13.11 (Nov. 2017), e1005783. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1005783.
- [31] Yonglin Wang et al. “Crash Prediction Using Deep Learning in a Disorienting Spaceflight Analog Balancing Task”. In: *Frontiers in Physiology* 13 (Jan. 2022), p. 806357. ISSN: 1664-042X. DOI: 10.3389/fphys.2022.806357.
- [32] Qingsong Wen et al. “Time Series Data Augmentation for Deep Learning: A Survey”. In: *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*. Aug. 2021, pp. 4653–4660. DOI: 10.24963/ijcai.2021/631. eprint: 2002.12478 (cs, eess, stat).
- [33] Irina Zgonnikova, Arkady Zgonnikov, and Shigeru Kanemoto. “Stick Must Fall: Using Machine Learning to Predict Human Error in Virtual Balancing Task”. In: *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*. Dec. 2016, pp. 173–177. DOI: 10.1109/ICDMW.2016.0032.