Using Learnable Physics for Real-Time Exercise Form Recommendations

Abhishek Jaiswal
abhi.jaiswal44@gmail.com
Indian Institute of Technology, Kanpur
Kanpur, Uttar Pradesh, India

Gautam Chauhan
gautamchauhan0412@gmail.com
Indian Institute of Technology, Kanpur
Kanpur, Uttar Pradesh, India

Nisheeth Srivastava
nisheeths@gmail.com
Indian Institute of Technology, Kanpur
Kanpur, Uttar Pradesh, India

ABSTRACT

Good posture and form are essential for safe and productive exercising. Even in gym settings, trainers may not be readily available for feedback. Rehabilitation therapies and fitness workouts can thus benefit from recommender systems that provide real-time evaluation. In this paper, we present an algorithmic pipeline that can diagnose problems in exercises technique and offer corrective recommendations, with high sensitivity and specificity, in real-time. We use MediaPipe for pose recognition, count repetitions using peak-prominence detection, and use a learnable physics simulator to track motion evolution for each exercise. A test video is diagnosed based on deviations from the prototypical learned motion using statistical learning. The system is evaluated on six full and upper body exercises. These real-time recommendations, counseled via low-cost equipment like smartphones, will allow exercisers to rectify potential mistakes making self-practice feasible while reducing the risk of workout injuries.

CCS CONCEPTS

• **Computing methodologies** → **Interactive simulation**; **Real-time simulation**; • **Human-centered computing** → *Human computer interaction (HCI).*

KEYWORDS

real-time exercise pose recommendations, physics-inspired neural networks

ACM Reference Format:
Abhishek Jaiswal, Gautam Chauhan, and Nisheeth Srivastava. 2023. Using Learnable Physics for Real-Time Exercise Form Recommendations. In *Seventeenth ACM Conference on Recommender Systems (RecSys '23), September 18–22, 2023, Singapore, Singapore.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3604915.3608816

1 INTRODUCTION

Sedentary lifestyles and physical inactivity are prominent risk factors for cardiovascular diseases worldwide. Evidence also suggests that physical activity has dipped considerably over time [21]. Exercise improves life expectancy and has an effective therapeutic impact on physical and mental health [13]. Assistance in performing physical exercises [10] , therefore, has a vital role to play in improving health on a global scale.

Expert supervision in performing exercises is not readily available and even if present, self practice needs further assistance. Consequently, digital technologies have become instrumental in improving accessibility to expert supervision. In gym settings, Sensor-based methods [26, 27] have been prolifically used for pose assessment focusing on exercise detection [6, 22, 23], rep counting [25, 26], incorrect pose diagnosis [12, 15, 16, 33] and suggestions [26, 27, 34]. However, sensors can be obtrusive, expensive, and difficult to calibrate correctly, and so may be best suited for high-performance settings [14]. More appropriate for less intensive settings, vision-based methods [11, 17, 30, 31] have recently gained prominence due to improvements in deep learning techniques and mobile camera technology. This direction of research is very promising because it allows for the possibility of entirely sensor-free tracking of exercise performance.

At present, however, such proposals face considerable difficulties. Most vision-based approaches to exercise tracking work with predetermined heuristic parameters which vary across exercises and participants, requiring considerable hand-crafting [7, 17]. While vision-based approaches to exercise type recognition and rep-counting are plentiful, approaches that seek to track exercise form are limited to simple upper body exercises with relatively little body movement [7, 15, 17]. Further, most such approaches offer exercise diagnoses retrospectively after processing entire recorded exercise sessions [14, 25].

Taking the conventional idea of recommendation from collaborative filtering a step further, we frame the task of exercise supervision as that of recommending correct forms. To this end, we identify the over-general nature of the deep learning architectures used in vision-based exercise tracking pipelines as a key problem blocking progress in this area. Rather than use generic neural network architectures, we propose using a specific variety of neural networks, specifically designed to learn relationships between physical objects, as the base inference engine in such recommender systems. Using one such architecture - Interaction Networks [2] - we describe a novel recommender system for real-time exercise form correction in this paper. We show that our solution works with very high sensitivity and specificity for a wide variety of full-body and upper-body exercises.

Section 2 places our proposal in line with existing approaches of exercise-specific recommendations. We explain the working of our physics inference model in Section 3 and the components of

RecSys '23, September 18–22, 2023, Singapore, Singapore

ACM ISBN 979-8-4007-0241-9/23/09…$15.00
https://doi.org/10.1145/3604915.3608816

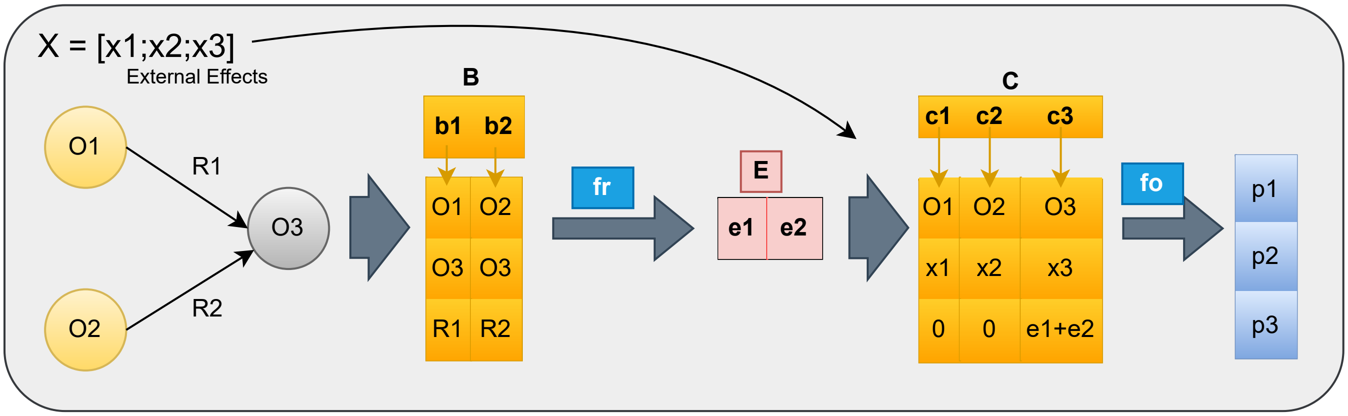

Figure 1: Example showing working of Interaction Network. Vector b includes sender and receiver object details, Vector e is the resulting effect of an interaction. Vector c includes interaction effects and the exercised object.

our recommender system in Section 4. The value of our system is demonstrated through experiments on different exercises in Section 5. Finally, we conclude by highlighting the salient points of our recommender system and identifying directions for future work, in Section 6.

2 RELATED WORK

Several recently proposed systems [20, 29, 31] have used state-of-the-art pose estimation techniques to craft heuristic joint angle thresholds for pose feedback. Recently, real-time pose diagnosis was done by Alatiah and Chen [1] using pre-calculated range of motion. Similarly, Ying et al. [32] used pre-stored correct exercises to detect incorrect moves. Such systems offer only binary feedback without any corrective recommendations.

More granular diagnoses are possible in a system recently proposed by Liu and Chu [17], who learned three joint angle indicators for each rep using a RNN and provided visual diagnosis for two simple upper body dumbbell exercises with high accuracy. However, their approach requires per frame annotation for training. Similarly, Gharasuie et al. [11] developed a low-cost system using AlphaPose[9] based arm angles for upper-body exercises to count reps and also quantify exercise phase parameters to estimate user fatigue levels. While such heuristic based methods provide helpful textual feedback in some instances, they tend to work well only for isolation arm exercises involving only few joints and do not achieve significant diagnostic accuracy without extensive frame-level annotation. Our system, in contrast, with a more sophisticated inference engine, works well for compound exercises using only video-level annotation.

Closer technically to our approach, Pose Trainer [7] uses OpenPose [4] on dumbbell exercises along with Dynamic Time Warping against template moves for rep diagnosis. They use angular heuristics for exercise feedback. Similarly, AI Coach [30] compares sports trajectories against pre-annotated bad poses, recommending an exemplar-based video for all identified bad pose frames. This is in contrast with our system, wherein holistic, real-time, body-focused textual feedback is offered to users, immediately enabling them to correct any unsafe exercise posture.

Table 1: Body Landmarks for four full body exercises.

Squats	Sit-ups	Push-ups	Lunges
Nose	Nose	Shoulder	Shoulder
Left Hip	Shoulder	Hand	Hip
Right Hip	Hip	Hip	Front Knee
Left Knee	Knee	Toe	Back Knee
Right Knee	-	-	Back Toe
-	-	-	Front Heel

3 PHYSICS PREDICTION USING INTERACTION NETWORK

Interaction Network (IN) [2] is a Graph Networks based general-purpose physics engine simulator. The graph nodes represent the landmark joints (Table 1), and the edges represent the natural joint relations. The IN (Figure 1) entails a relation-centric function f_R for predicting interaction effects and an object-centric function f_O for predicting the next step dynamics using the calculated interaction effects. It feeds the sender landmarks' properties(O_S), the receiver landmark properties(O_R), and their relational properties(R_a) as matrices at the current time to the relation-centric function f_R, which outputs the effect matrix E, therefore we have

$$E = f_R(O_S; O_R; R_a) \quad \textit{where ; means concatenation.}$$

Product of the effect matrix with the binary receiver matrix R_r yields $\bar{E} = ER_r^T$ assimilating the net effects on each receiver object in its columns. This, along with the object matrix O, is fed to the object-centric function f_O to predict the next state P of each landmark.

$$P = f_O(O; \bar{E})$$

For more information, readers are requested to refer to the IN paper [2].

4 PROPOSING POSE CORRECTING RECOMMENDATIONS

We first outline the overall methodology of our pipeline, followed by a detailed description of its sub-components. To begin with, we feed

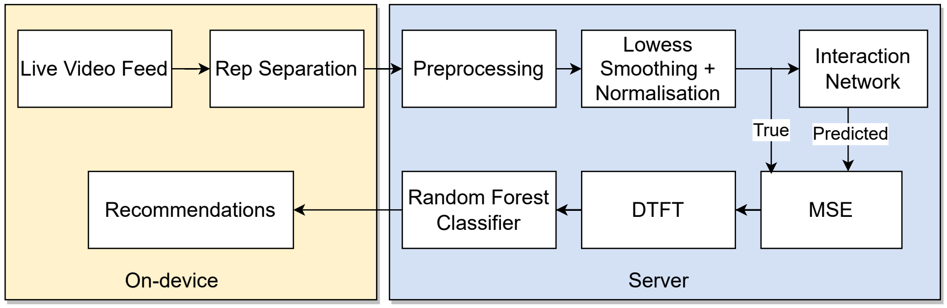

Figure 2: Flowchart illustrating our Interactive System's workflow.

a recorded or live video to our pipeline, which predicts per-frame keypoints for 25 joints through Mediapipe API [3]. Depending on motion evolution, we select exercise-specific landmarks (Table 1) followed by normalization and smoothing for physics modeling. The ML model predicts the motion rollouts for all the landmarks with visibility of only the initial rep state. Using these predictions, we calculate the Mean Squared Error (MSE) for individual landmarks and then transform them to the frequency domain for further processing, as described in subsequent sections. Essentially, we use frequency domain information from the MSE signals to classify exercise reps as either correct or incorrect (in one of the multiple predefined modes of failure) using a Random Forest multi-class classifier (Figure 2). Thus, our pipeline receives visual input from the user side and emits textual recommendations from the model side.

4.1 Rep Counting using Peak Prominence

We exploit cyclic movements within each exercise for rep segregation. To that end, we find peaks in the periodic landmark displacement plot. These peaks may contain extraneous motion data, such as fragments between successive reps and other discontinuities from tired and distracted performers. To detect genuine peaks, we find peaks' importance using peak prominence and use its standard deviation as a cutoff for our high pass filter. Displacement values above the cutoff delineate the start and stop of a valid rep.Figure 3 shows the result of rep counting for a single lunges video.

4.2 Preprocessing

MediaPipe API provides 3D positional time series data for 25 body landmarks for each exercise. We transform these coordinates for unidirectional facing and use the resulting view along with the landmark's displacement amplitude to fix the representative landmarks for an exercise.

We apply Locally Weighted Scatterplot Smoothing (LOWESS) to each time series [8] to reduce noise, discard reps with significant pose estimation errors and Min-Max normalize the coordinates to induce translational invariance, finally outputting a stick figure representations for each rep (Figure 4).

4.3 Learning exercise dynamics

We represent body landmarks as graph nodes and the natural body connections as the edges, one each for the forward and backward direction. The choice of landmarks depended on our understanding of the biomechanics of each exercise. We consider only 2D position and velocity as nodes' attributes. Relational attributes include joint-to-joint distances and angles, with positive and negative unity indicating edge direction. A top and a bottom stationary reference point is added to graph nodes with x coordinates 0.5 and y values as 0 and 1. Velocity is approximated as the difference between current and previous coordinates. The two feed-forward neural networks - the relation-centric model f_R and the object-centric model f_O consist of three and four hidden layers respectively with each layer having 256 units with ReLU activation and dropout value 0.5. The output layers have linear activation. We train the model for 2500 epochs with early stopping using AdamW optimizer [18] with 1cycle learning rate policy [24]. and a learning rate of 3e-4.

Table 2: Recommendations offered for full body exercises.

Lunges
Keep your knees behind the toes
Keep your legs closer, they are too wide apart
Squats
Keep your knees behind the toes
Don't bend your knees inward
Keep your Feet shoulder-width apart
Situps
Your back should rise up completely
Pushups
Keep your Knees-hips-Shoulders in a straight line
Lower your chest to align it with hip
Lower your hips
Your chest should not touch the ground

Table 3: F1 scores of training with different classifiers.

Classifier	Squats	Push-ups	Lunges	Sit-ups
SVM	0.81	0.96	0.93	0.97
KNN	0.75	0.87	0.94	0.92
Naive Bayes	0.74	0.95	0.92	0.98
Logistic Reg.	0.85	0.97	0.95	0.98
LDA	0.87	0.98	0.97	0.97
Random Forest	**0.94**	**0.98**	**0.97**	**0.98**

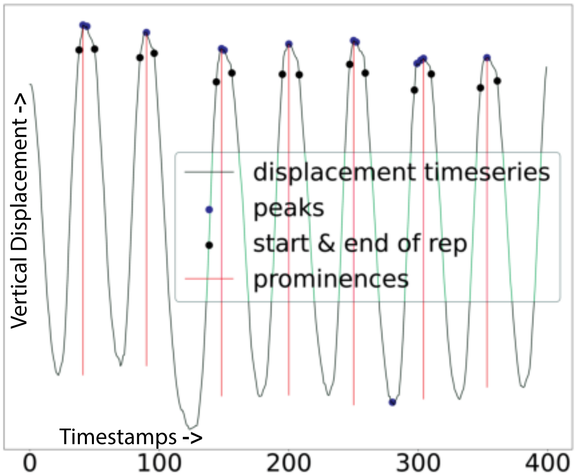

Figure 3: Peak prominence over vertical periodicity for rep counting.

Figure 4: Push-ups - stick figure and corresponding video frame.

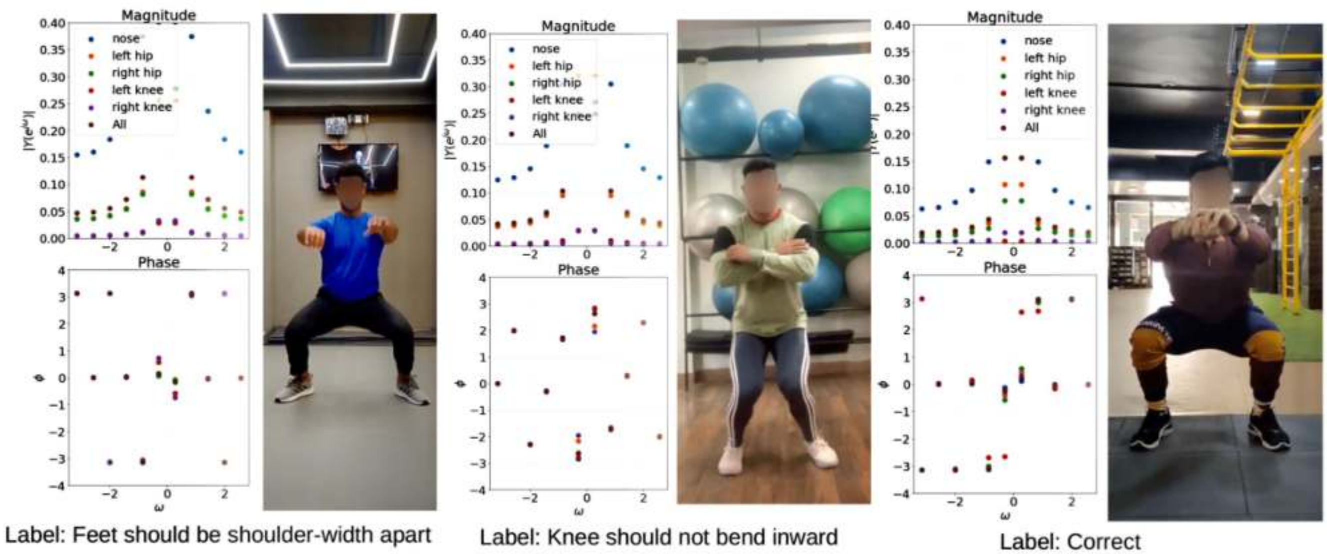

Figure 5: DTFT representations of error signature of different squats labels. Note how the phase plots for the incorrect squats differ from each other as well as from a correct squat systematically.

4.4 Error Analysis and Rep Classification

The MSE time series emitted by the IN informs the subsequent stages. We hypothesize that our physics engine predicts the correct method of exercising, such that considerable deviation from it would hint at an incorrect rep. Further, the specific combination of MSE from different body components would hint at the precise mistake made by the exerciser.

To extract this information, we transform the MSE time series from all the representative landmarks to the frequency domain using the discrete-time Fourier transform (DTFT). DTFT provides magnitude and phase values for each time series. Conversion to the frequency domain helps in two ways. It gives a fixed-sized representation of the variable-length time series. It also helps to extract the features of the time series. This output of the DTFT, called the error signature (Figure 5), is a vector representation of an exercise rep of variable duration. For our case, we take the principal 11 amplitudes and the corresponding phase values to build the error signature of each exercise rep. At the final stage of our pipeline, we use a Random Forest classifier for classification, operated in a multi-class classification setting, with recommendation labels used as shown in Table 2. We found Random Forest was the most consistent classifier across all the exercises. (Table 3) and tuned its hyperparameters using randomized search cross-validation

Table 4: Baselines comparisons for four full body exercises (up) and two upper body exercises(down). Classification results reported using weighted F1 score with standard deviations over five train-test runs (*Shoulder Press results reported for two incorrect classes).

Model	Squats	Push-ups	Lunges	Sit-ups
MLP	0.91 ± 0.02	0.98 ± 0.03	0.95 ± 0.03	**0.99 ± 0.01**
RNN	0.85 ± 0.04	0.98 ± 0.01	0.94 ± 0.01	0.98 ± 0.02
GRU	0.87 ± 0.03	0.98 ± 0.01	0.93 ± 0.02	0.94 ± 0.04
IN	**0.94 ± 0.02**	**0.98 ± 0.01**	**0.97 ± 0.01**	0.98 ± 0.01

Model	ShoulderPress*	FrontRaise
Ng [20]	0.90	0.77
PoseTrainer[7]	0.49	0.76
MLP	**0.99 ± 0.01**	0.82 ± 0.04
RNN	0.99 ± .01	0.79 ± .05
GRU	0.95 ± .06	0.80 ± .04
IN	0.98 ± 0.01	**0.88 ± 0.03**

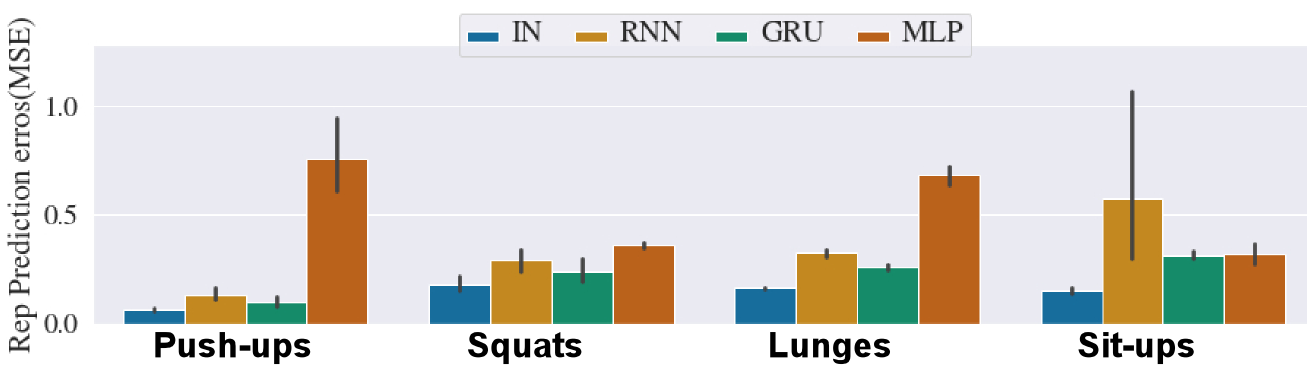

Figure 6: Average rollout prediction errors over exercise reps(MSE) for Baseline Models and Interaction Network. Even though MLP have good F1 scores (Table 4) in some cases, high prediction error makes their performance unreliable.

5 EMPIRICAL EVALUATION

5.1 Data

For the evaluation of full-body exercises, we used a proprietary dataset from E-Trainer Analytics Wizard Pvt. Ltd. This dataset contains the front and side view of seven exercisers performing more than 150 reps for four exercises - squats, push-ups, lunges, and sit-ups. Each exercise has one correct class, whereas incorrectly done exercises could belong to multiple classes. Incorrect videos were annotated with corrective recommendations by expert physical trainers.

We used a train test split of 60%-40% to train the classifier on incorrect classes. Similarly, 60% of the correct class data was used to train our physics engine. To compare against existing approaches of form prediction, we conducted evaluation using annotated data for shoulder press and front raise from a publicly available dataset [20]

5.2 Baseline Comparisons

Among available learnable dynamics predictors, we exploit IN for their interpretability and simplicity. Our pipeline can also function with other motion predictors to the degree that they can accurately mimic the dynamics of the exercise. To test this hypothesis, we evaluate our model against several baselines.

The **Multi layer perceptron (MLP) Baseline**, with three 256-length hidden layers and ReLU activation, has all information to learn the interaction dynamics, requiring it to assimilate relation indices implicitly without explicit scene factorization. The **Recurrent Neural Network (RNN) and Gated Recurrent Unit (GRU)**, capable of modeling posture evolution, have three recurrent units with three features in the hidden state. The final hidden state output is fed to a fully connected layer to predict future dynamics. All the above baselines use flattened node attributes as input.

Additionally, we compare our pipeline against popular heuristic techniques ([20], [7]), which examine reps using geometric thresholds over joint features. Next, we describe ablation modifications on IN's architecture and input.

The **Attribute Hidden IN** uses the same IN architecture but with an empty relation attribute matrix which, in principle, could be deduced from position data demanding estimation of complex distance and inclination functions. The **Independent object IN** simulates removing the relation-centric component by zeroing out

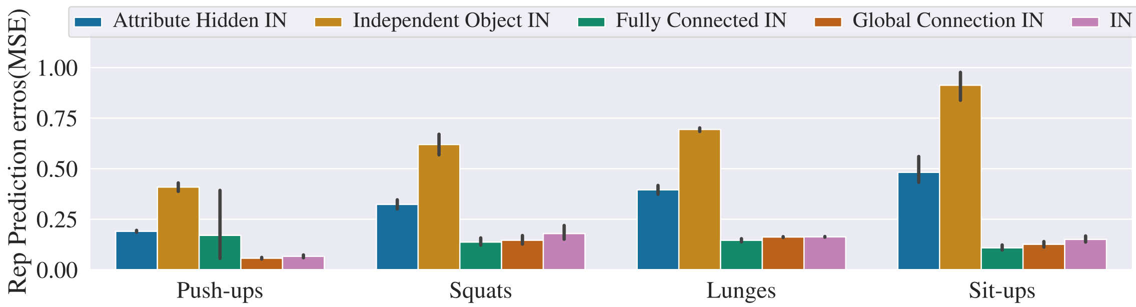

Figure 7: Average rollout prediction error over exercise reps(MSE) for Ablation Models and Interaction Network. Models without relation information experience significant drop in performance for dynamics prediction.

the interaction effects, which incapacitates it from modeling object-object interactions. The **Fully Connected (FC) IN** connects each joint with every other joint, procuring the same capacity as the IN but involving additional irrelevant inputs. The **Global Connection (GC) IN**, additionally, connects all the landmark points to the two stationary reference points, modeling both local and global interactions for superior information propagation.

5.3 Results

We evaluate our diagnostic system on six exercises using three criteria: Rep Counting , Posture diagnosis, and Real-time prediction. Our peak-prominence based algorithm perfectly counted all the reps in the full-body exercises. For each rep detected, we measured recommendation accuracy using weighted F1 scores in a multi-class classification setting.

5.3.1 Posture diagnosis. Our results (Table 4) indicate that a pipeline endowed with physics learning capability outperforms all baselines effectively differentiating correct and incorrect exercise reps. Performance can be comparable for fewer prediction classes (e.g., Sit-ups, Shoulder Press) or simple exercises with a small range of joint motion(e.g., Push-ups). However, the performance gain is evident as the number of incorrect classes increases (Front Raise - Table 5).

Since classification results are an indirect measure of physics modeling, we also explore the MSE for the next state prediction (Figure 6). In all cases, a physics learning engine best describes motion dynamics. Even though the MLP showed good classification (Table 4), it significantly deviates from the actual exercise, causing deviant performance, especially with increasing exercise complexity (Table 5). Even an ill-suited model can classify well if the error signatures are separable. However, such arbitrary performance gains do not scale well as the number of classes increases.

Our ablation study on IN's performance gain indicates that relational attributes are critical for learning the physics of interactions (Figure 7). Reasonable variations in the joint-to-joint links (as with FC-IN and GC-IN) show similar effects, with GC-IN edging slightly over the vanilla IN, probably due to faster information propagation. Stochastic interactions with involuntary factors (like fatigue and distractions) possibly contain the expected performance gain from global propagation. The FC-IN with redundant information in its irrelevant relations competes well, presumably because the IN learns to weigh the importance of exercise-specific relations. This architecture may, thus, bypass the requirement of an explicit relational matrix, provided that the pose estimation accurately detects all body landmarks. As expected, the independent object IN without relation function f_r finds it challenging to model the motion dynamics. Similarly, the Attribute Hidden IN is unable to exploit the joints' information without relational attributes.

Table 5: Baselines Comparison for Front Raise with increasing number of classes . Methods with poor dynamics modeling show significant performance drop.

FrontRaise	2 Classes	4 Classes	6 Classes
MLP	0.96	0.91	0.82
RNN	0.93	0.90	0.79
GRU	0.93	0.89	0.80
IN	0.96	0.91	0.88

Table 6: Lag time(seconds) for new rep recognition.

Exercise	Mean(sec)	Standard deviation(sec)
Squats	0.55	0.13
Sit-ups	0.39	0.07
Push-ups	0.36	0.11
Lunges	0.54	0.09

5.3.2 Diagnosis latency . The MediaPipe Android API, fed with the exercise's camera feed, outputs the joints' coordinates time series. After rep segregation, each rep data is passed to the server, where our pipeline classifies it as correct or diagnoses it as a mistake of a particular type. A corrective recommendation specific to the estimated diagnosis is displayed to the user through our mobile application. For exercisers operating at normal tempo, this feedback arrives before their next rep is halfway complete prompting the user to instantly correct any mistakes in technique (see Table 6 for a quantitative summary and Table 7 for examples of live recommendations videos).

6 DISCUSSION AND LIMITATIONS

Self-training is gaining popularity as more and more people feel disinclined to commit to a dedicated gym routine. By providing accurate real-time recommendations, we can benefit such users without

Table 7: Link to videos of interaction network predictions rolled out over time and of exercise sessions diagnosed using our system in real-time.

Data type	Link
IN predictions simulations	Link
Exercise Demo	Link

compromising their day-to-day schedule. Our recommender system focuses on rep counting and diagnosis, assuming that the exercise performed is known (or is easily knowable). For instance, Moran et al. [19] used MediaPipe API [3] for pose recognition to detect the type of exercise performed in real-time, a capability that could easily inform exercise type in our pipeline.

Thus, to summarize, this paper offers a novel system for recommending form corrections to exercisers performing rep-based training in real-time with high precision. We introduce the use of learnable physics engines to model body physics, a task for which they are very well-suited. The success of our physics model permits downstream classifiers to accurately diagnose modes of failure of exercises using differential prediction error residuals between the model prediction and actual observations. Empirical evaluations show that our system diagnoses defective techniques in complex full-body exercises with high sensitivity and specificity. We expect the adoption of such interactive systems to help healthcare providers scale up access to supervised physical exercise.

We conclude with a brief exploration of the limitations of our system, and possible directions for future work. The most critical technical limitation of the present system is its reliance on pre-defined relational attributes for each exercise's Interaction Network. These attributes depend on the nature of human biomechanics and must be decided beforehand. Learning relational attributes from data could improve this performance even further, a clear direction for future work. Our system is currently tested only for exercises with significant vertical periodicity, an artifact of our peak-prominence based rep-counting scheme, though vertical periodicity also exists in many other exercises. Replacing this with a more sophisticated rep-counting method could extend our system's capabilities to a more general set of exercises. In particular, given the known diagnostic value of gait analysis in predicting health outcomes for the elderly [5, 28], extending this system's digital diagnostic capabilities to monitoring and diagnosing gait-related problems presents a very promising direction for future work.

REFERENCES

[1] Talal Alatiah and Chen Chen. 2020. Recognizing exercises and counting repetitions in real time. *arXiv preprint arXiv:2005.03194* (2020).

[2] Peter Battaglia, Razvan Pascanu, Matthew Lai, Danilo Jimenez Rezende, et al. 2016. Interaction networks for learning about objects, relations and physics. *Advances in neural information processing systems* 29 (2016).

[3] Valentin Bazarevsky, Ivan Grishchenko, Karthik Raveendran, Tyler Zhu, Fan Zhang, and Matthias Grundmann. 2020. Blazepose: On-device real-time body pose tracking. *arXiv preprint arXiv:2006.10204* (2020).

[4] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2021. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 43, 01 (2021), 172–186.

[5] Matteo Cesari, Stephen B Kritchevsky, Brenda WHJ Penninx, Barbara J Nicklas, Eleanor M Simonsick, Anne B Newman, Frances A Tylavsky, Jennifer S Brach, Suzanne Satterfield, Douglas C Bauer, et al. 2005. Prognostic value of usual gait speed in well-functioning older people—results from the Health, Aging and Body Composition Study. *Journal of the American Geriatrics Society* 53, 10 (2005), 1675–1680.

[6] Keng-hao Chang, Mike Y Chen, and John Canny. 2007. Tracking free-weight exercises. In *International Conference on Ubiquitous Computing*. Springer, 19–37.

[7] Steven Chen and Richard R Yang. 2020. Pose Trainer: correcting exercise posture using pose estimation. *arXiv preprint arXiv:2006.11718* (2020).

[8] William S Cleveland. 1981. LOWESS: A program for smoothing scatterplots by robust locally weighted regression. *American Statistician* 35, 1 (1981), 54.

[9] Hao-Shu Fang, Shuqin Xie, Yu-Wing Tai, and Cewu Lu. 2017. Rmpe: Regional multi-person pose estimation. In *Proceedings of the IEEE international conference on computer vision*. 2334–2343.

[10] Gerald F Fletcher, Carolyn Landolfo, Josef Niebauer, Cemal Ozemek, Ross Arena, and Carl J Lavie. 2018. Promoting physical activity and exercise: JACC health promotion series. *Journal of the American College of Cardiology* 72, 14 (2018), 1622–1639.

[11] Morteza Mohammady Gharasuie, Noah Jennings, and Shubham Jain. 2021. Performance Monitoring for Exercise Movements using Mobile Cameras. In *Proceedings of the Workshop on Body-Centric Computing Systems*. 1–6.

[12] Oonagh M Giggins, Kevin T Sweeney, and Brian Caulfield. 2014. Rehabilitation exercise assessment using inertial sensors: a cross-sectional analytical study. *Journal of neuroengineering and rehabilitation* 11, 1 (2014), 1–10.

[13] David Jiménez-Pavón, Ana Carbonell-Baeza, and Carl J Lavie. 2020. Physical exercise as therapy to fight against the mental and physical consequences of COVID-19 quarantine: Special focus in older people. *Progress in cardiovascular diseases* 63, 3 (2020), 386.

[14] Rushil Khurana, Karan Ahuja, Zac Yu, Jennifer Mankoff, Chris Harrison, and Mayank Goel. 2018. GymCam: Detecting, recognizing and tracking simultaneous exercises in unconstrained scenes. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (2018), 1–17.

[15] Yousef Kowsar, Masud Moshtaghi, Eduardo Velloso, Lars Kulik, and Christopher Leckie. 2016. Detecting unseen anomalies in weight training exercises. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction*. 517–526.

[16] Jaehyun Lee, Hyosung Joo, Junglyeon Lee, and Youngjoon Chee. 2020. Automatic classification of squat posture using inertial sensors: Deep learning approach. *Sensors* 20, 2 (2020), 361.

[17] An-Lun Liu and Wei-Ta Chu. 2020. A posture evaluation system for fitness videos based on recurrent neural network. In *2020 International Symposium on Computer, Consumer and Control (IS3C)*. IEEE, 185–188.

[18] Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017).

[19] Alex Moran, Bart Gebka, Joshua Goldshteyn, Autumn Beyer, Nathan Johnson, and Alexander Neuwirth. 2022. Muscle Vision: Real Time Keypoint Based Pose Classification of Physical Exercises. *arXiv preprint arXiv:2203.12111* (2022).

[20] Jiunn Ng. 2020. *Posture evaluation for variants of weight-lifting workouts recognition*. Ph. D. Dissertation. UTAR.

[21] Cemal Ozemek, Carl J Lavie, and Øivind Rognmo. 2019. Global physical activity levels-Need for intervention. *Progress in cardiovascular diseases* 62, 2 (2019), 102–107.

[22] Christian Seeger, Alejandro P Buchmann, and Kristof Van Laerhoven. 2011. myHealthAssistant: a phone-based body sensor network that captures the wearer's exercises throughout the day.. In *BodyNets*. 1–7.

[23] Goran Šeketa, Dominik Džaja, Sara Žulj, Luka Celić, Igor Lacković, and Ratko Magjarević. 2015. Real-time evaluation of repetitive physical exercise using orientation estimation from inertial and magnetic sensors. In *First European Biomedical Engineering Conference for Young Investigators*. Springer, 11–15.

[24] Leslie N Smith. 2018. A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay. *arXiv preprint arXiv:1803.09820* (2018).

[25] Andrea Soro, Gino Brunner, Simon Tanner, and Roger Wattenhofer. 2019. Recognition and repetition counting for complex physical exercises with deep learning. *Sensors* 19, 3 (2019), 714.

[26] Gabriele Spina, Guannan Huang, Anouk Vaes, Martijn Spruit, and Oliver Amft. 2013. COPDTrainer: a smartphone-based motion rehabilitation training system with real-time acoustic feedback. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. 597–606.

[27] Eduardo Velloso, Andreas Bulling, Hans Gellersen, Wallace Ugulino, and Hugo Fuks. 2013. Qualitative activity recognition of weight lifting exercises. In *Proceedings of the 4th Augmented Human International Conference*. 116–123.

[28] Joe Verghese, Roee Holtzer, Richard B Lipton, and Cuiling Wang. 2009. Quantitative gait markers and incident fall risk in older adults. *The Journals of Gerontology: Series A* 64, 8 (2009), 896–901.

[29] Parth Vyas. 2019. Pose estimation and action recognition in sports and fitness. (2019).

[30] Jianbo Wang, Kai Qiu, Houwen Peng, Jianlong Fu, and Jianke Zhu. 2019. Ai coach: Deep human pose estimation and analysis for personalized athletic training

assistance. In *Proceedings of the 27th ACM International Conference on Multimedia*. 374–382.

[31] Lehao Wang, Yucheng Chen, and Weijun Duan. 2021. Monocular Keypoint based Pull-ups Measurement on Strict Pull-ups Benchmark. In *2021 4th International Conference on Computer Science and Software Engineering (CSSE 2021)*. 307–311.

[32] Haocong Ying, Tie Liu, Mingxin Ai, Jiali Ding, and Yuanyuan Shang. 2021. AICoacher: A System Framework for Online Realtime Workout Coach. In *Proceedings of the 29th ACM International Conference on Multimedia*. 3787–3790.

[33] Aras Yurtman and Billur Barshan. 2014. Automated evaluation of physical therapy exercises using multi-template dynamic time warping on wearable sensor signals. *Computer methods and programs in biomedicine* 117, 2 (2014), 189–207.

[34] Wenbing Zhao, Roanna Lun, Deborah D Espy, and M Ann Reinthal. 2014. Realtime motion assessment for rehabilitation exercises: Integration of kinematic modeling with fuzzy inference. *Journal of Artificial Intelligence and Soft Computing Research* 4, 4 (2014), 267–285.