Deep Learning for Squat Phase Classification in Powerlifting: LSTM Approach VS Traditional ML Approach

Fort Hunter, Blake Milstead, Doga Topcicek, Linda Ungerboeck, Ryan Franqui
COSC 424/525
The University of Tennessee
Emails: {lhunte21, bmilstea, dtopcice, lungerbo, rfranqui}@vols.utk.edu

***Abstract*—We present a computer vision system for classifying squat movement phases in powerlifting. Using Mediapipe's Pose Model, we extract 3D skeletal data (33 keypoints at 0.3s intervals) from Olympic weightlifting videos. Our approach evaluates four machine learning classifiers (logistic regression, ridge classification, random forest, and gradient boosting) to distinguish concentric from eccentric movements, with optimization for recall to minimize false negatives. We address dataset challenges including spotter interference through video preprocessing techniques. The current binary classification framework will be extended to identify specific form errors (spinal alignment, knee valgus, and foot positioning) in future work. This system provides an accessible alternative to professional coaching for technique correction in powerlifting training.**

I. Introduction

The use of computer vision to analyze and correct athletic movements has been a topic of interest for researchers and practitioners alike, with applications ranging from injury prevention to performance enhancement. The importance of proper technique in powerlifting exercises cannot be overstated, as improper form can lead to serious injuries and long-term damage. Despite the growing recognition of this issue, many athletes struggle to identify weak points in their technique due to a lack of immediate feedback or access to expert coaching.

Current solutions for addressing these issues often rely on personal training sessions with experienced coaches, which can be expensive and inaccessible to those who cannot afford them. Furthermore, even trained professionals may find it challenging to pinpoint subtle errors in an athlete's form without extensive practice and experience. The need for a cost-effective solution that provides real-time feedback on posture correction has become increasingly pressing.

In response to this challenge, our team has developed a novel approach using computer vision techniques to analyze videos of powerlifting exercises. By estimating joint data from video analysis, we can provide immediate feedback on areas that require improvement, enabling athletes to adjust their technique without relying on external guidance. This technology offers a game-changing solution for both beginners and professionals seeking to optimize their training while minimizing the risk of injury.

The goal of this work is to construct an accurate classifier capable of distinguishing between concentric and eccentric movements in squat exercises. By achieving this objective, we aim to provide athletes with a valuable tool that can be integrated into their existing training regimens, promoting safer and more effective practice without breaking the bank. The rest of this paper outlines our technical approach, baseline selection, implementation, and future extensions of our network.

II. Literature Review

Human pose estimation plays an important role in analyzing athletic movements, particularly in the context of powerlifting, where proper technique is crucial for both performance optimization and injury prevention. Recent advancements in computer vision have significantly improved the ability to analyze human posture and movement during exercises such as squats.

One method for pose estimation is Mediapipe, a framework that utilizes machine learning models to extract 33 key joint landmarks from the video data. This framework enables real-time 3D skeletal representation, which can be utilized to assess and correct posture during exercises (Liu et al., 2023). Mediapipe's Pose Model offers high accuracy and efficiency, making it ideal for applications in sports performance analysis and injury prevention.

In terms of classifying movements, previous research has focused on using machine learning models to distinguish between different phases of an exercise. For example, the use of random forests and gradient boosting classifiers has proven successful in analyzing dynamic movements in powerlifting exercises. These classifiers are particularly useful due to their ability to handle complex, nonlinear relationships inherent in human posture data (Zhao et al., 2023). Gradient boosting methods, in particular, offer strong performance for detecting subtle variations in joint movement, which is essential for accurately identifying concentric and eccentric phases during squat exercises.

Deep learning models, especially recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, have also shown promise in recognizing time-dependent activities such as squats. These models excel in learning from sequential data, making them ideal for tasks that require tracking movement over time (Yin et al., 2023). Incorporating LSTMs into powerlifting pose analysis has the potential to improve the accuracy of movement phase classification by leveraging temporal dependencies between key body points during the squat.

Furthermore, recent studies have explored the integration of skeletal data and video analysis techniques for automatic correction of exercise technique. These approaches could provide athletes with real-time feedback on their performance, ultimately improving training outcomes and reducing the risk of injury.

In summary, the use of advanced pose estimation and machine learning techniques holds great promise for optimizing powerlifting techniques and providing athletes with valuable feedback for improving their performance.

III. Dataset

The data for this project was collected from the Olympic Powerlifting Championships. We gathered videos of professionals across various heights, weights, and genders to improve the breadth of our model. Each video captures a single repetition of a correctly executed squat. Once we obtained all of these videos, we converted them into a 3D skeletal representation using Mediapipe's Pose Model. This representation contained 33 unique spatial body key points, represented as x, y, and z, and collected every third of a second.

After creating the dataset, we noticed that the model was having some trouble staying focused on the individual squatting. Since this is an event that has professionals lifting huge amounts of weight, there are spotters in the background that are meant to assist the lifters if they are unable to complete the lift. These spotters added some complications to our data. Oftentimes during the lifts, our model would switch from tracking the lifter to tracking the spotter. While we haven't been able to completely get rid of this issue yet, we have come up with some solutions that have helped. Our first solution was to crop the videos in a way in which we reduce the amount of the spotter's body that is able to be seen. This helped quite a bit, especially with the spotters that are to the left and right of the squatter. However, this does not help with the spotter that stands behind the squatter. Since the spotter and the squatter occupy the same part of the frame, it is impossible to remove them by cropping. This led us to our next solution: editing our videos to remove the spotter completely. While this is laborious and we have not been able to do this with all of our videos yet, it has yielded strong results. In the future we hope to do this with more of our videos to further improve our data.

To process the skeletal data into usable features for our classification models, we implemented a custom PyTorch Dataset class that performs several key transformations:

- **Biomechanical Angle Calculation:** We compute joint angles from the 3D coordinates, focusing on key biomechanical markers including knee flexion, hip flexion, and torso lean. The calculations incorporate physical constraints to ensure anatomically plausible ranges (e.g., knee angles constrained between 60-180 degrees).
- **Temporal Smoothing:** We apply Gaussian smoothing with feature-specific parameters ($\sigma = 2.0$ with increasing smoothing for higher-index features) to reduce noise while preserving important movement patterns.
- **Velocity Analysis:** The system calculates angular velocity from knee angle changes to help distinguish between movement phases, using a minimum velocity threshold (0.025 rad/frame) to filter insignificant movements.
- **Adaptive Sequence Labeling:** The dataset automatically labels sequences (30 frames with 66
  - **DOWN (0):** When mean velocity is negative and knee angle decreases significantly
  - **UP (1):** When velocity exceeds threshold and knee angle increases
  - **STABLE (2):** For transitional periods with minimal movement
- **Normalization:** We standardize features while preserving joint relationships through z-score normalization, storing original statistics for reference.
- **Balanced Sampling:** The DataLoader uses weighted random sampling to address class imbalance, ensuring equal representation of all movement phases during training.

The dataset pipeline includes comprehensive analysis tools that provide statistics on class distribution and feature characteristics. This processing transforms raw skeletal coordinates into meaningful biomechanical features that effectively capture the dynamics of squat movements for our classification task.

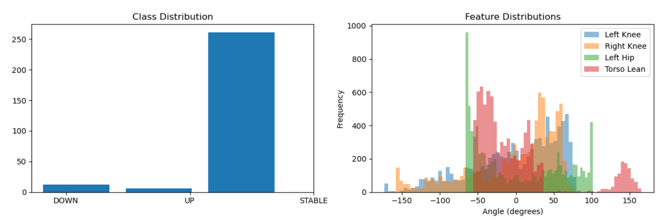

Fig. I: Analysis of the squat phase dataset showing (left) class distribution across movement phases (DOWN, UP, STABLE) and (right) feature distributions of key joint angles. The left plot shows the number of samples in each movement phase, while the right plot displays the frequency distribution of joint angles in degrees for left knee (blue), right knee (orange), left hip (green), and torso lean (red).

IV. Baseline Solution

To establish a robust foundation for our posture classification model, we employed a multi-algorithm approach to evaluate various machine learning (ML) techniques. Our primary objective was to develop an accurate classifier that can distinguish between eccentric and concentric movements in squat exercises.

The dataset used for this study consisted of video recordings of powerlifting exercises, annotated with labels indicating whether the movement was eccentric (downward motion) or concentric (upward motion). To create a baseline solution, we utilized four ML algorithms: logistic regression, ridge classification, random forest, and gradient boosting. These models were chosen based on their characteristics and suitability for posture classification tasks.

- **Logistic Regression:** Serves as a simple and interpretable model that provides a performance benchmark against which more complex models can be compared. It is widely used in binary classification tasks and allows for probability estimation of class membership.

- **Ridge Classification:** An extension of logistic regression that incorporates L2 regularization to prevent overfitting. This model is particularly useful given the potential for colinearity in our skeletal feature set.

- **Random Forest:** A robust ensemble learning technique that is effective in handling noisy data and complex decision boundaries. Given that posture classification involves nonlinear relationships between joint movements, random forest can capture these patterns efficiently.

- **Gradient Boosting:** A powerful boosting method that incrementally improves upon classification errors by combining weak learners. This model is expected to provide strong performance, particularly in cases where small variations in joint positioning significantly impact classification accuracy.

To assess the performance of these models, we plan to evaluate them based on precision, recall, and the F1-score, with a particular focus on recall. Since misclassification of movement phases can mislead athletes and potentially result in improper form correction, optimizing for recall ensures that we minimize false negatives in classification. In our literature review, random forest has demonstrated strong performance for movement classification tasks, but as we have not completed our data labeling process, we have not yet determined which approach will yield the highest accuracy for our specific dataset.

To complement these traditional machine learning approaches, we have also developed a Long Short-Term Memory (LSTM) network to leverage the temporal nature of our sequential movement data. The LSTM approach offers several theoretical advantages compared to traditional ML methods:

- **Temporal Processing:** Unlike static ML models that process each frame independently, LSTMs can learn patterns across sequences of frames (30-frame windows in our implementation), potentially capturing the dynamic transitions between movement phases.
- **Automatic Feature Learning:** While traditional ML requires careful feature engineering (e.g., calculating joint angles and velocities), the LSTM can learn relevant temporal features directly from raw joint position data.
- **State Maintenance:** The LSTM's memory cells can theoretically maintain information about previous phases to better interpret ambiguous intermediate positions.

However, the LSTM approach also presents certain challenges compared to traditional ML:

- **Data Requirements:** LSTMs typically require larger training datasets compared to traditional ML models to achieve good generalization.
- **Computational Cost:** The LSTM model involves significantly more parameters and requires more computational resources for training and inference.
- **Interpretability:** While traditional ML models (especially logistic regression) provide clear feature importance metrics, LSTM decision processes are more opaque.

Additionally, as we refine our baseline model, we will consider augmenting our feature set by incorporating temporal data, such as velocity and acceleration of key joints, to enhance the predictive capability of our classifiers. Future iterations of our approach may also involve deep learning-based techniques such as recurrent neural networks (RNNs) or convolutional neural networks (CNNs) to further improve classification accuracy.

This comprehensive evaluation framework allows us to compare both the interpretability and efficiency of traditional machine learning approaches with the temporal modeling capabilities of deep learning, providing a robust foundation for selecting the most appropriate technique for our movement classification task once experimental results are available.

V. Proposed Extension

The current objective of this research is to accurately classify eccentric and concentric movements in squats as either correct or incorrect. While this approach provides fundamental feedback on squat performance, it does not offer specific insights into what aspects of the movement are incorrect. In a more comprehensive approach, we aim to extend our classification system beyond a simple binary categorization.

Future iterations of our model will include additional classification labels that account for common squat form deficiencies. These refinements will enable the model to diagnose specific errors, providing athletes with actionable insights for improvement. Some of the key movement errors we plan to incorporate include:

- **Non-neutral spine positioning:** Detecting when an athlete fails to maintain a straight back during the squat.

- **Knees caving inward (valgus collapse):** Identifying improper knee alignment, which can indicate weakness or poor motor control.

- **Feet positioned too widely:** Recognizing when an athlete's stance is excessively wide, which may compromise stability and squat depth.

By integrating these error classifications, we will create a total of four distinct labels for both the eccentric (downward) and concentric (upward) phases of the squat: correct_up, non-neutral_spine, knees_caving, and feet_too_wide. This expanded classification system will enable more detailed feedback, allowing athletes to pinpoint specific deficiencies in their technique and make necessary corrections.

Beyond these enhancements, future work will explore the application of deep learning models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to improve classification accuracy. Additionally, incorporating real-time feedback mechanisms, possibly through wearable sensors or augmented reality interfaces, could further enhance the usability and effectiveness of our approach. By refining and expanding our classification framework, we aim to develop a highly accurate, accessible, and cost-effective tool for improving squat mechanics and reducing the risk of injury.

VI. Conclusion

In this project, we developed a pipeline in computer vision for analyzing squat technique using 3D pose estimation and machine learning classification. By using Mediapipe to extract joint key points from video data, we were able to begin distinguishing between concentric and eccentric squat phases with early results that are promising. For our future work, the goal is to expand this system to detect specific form errors during squats, enabling real-time feedback and safer, more effective training for athletes at all levels.

VII. Distribution Of Work

While we worked collaboratively on the project we all focused on certain parts of the project. Ryan worked on the literature review, looking into the previous related works, communicating with our mentor Dr. Santos, and collecting data. Doga worked on the baseline and LSTM model and helped with writing the midterm report. Linda helped with the baseline and labeling and helped with writing the midterm report. Fort helped by collecting and cleaning data, writing the midterm report, and collaborating on the model.

AI Disclosure

The author(s) would like to acknowledge the use of **ChatGPT**, a language model developed by **OpenAI**, in the preparation of this assignment. ChatGPT was used for **brainstorming ideas, improving grammatical clarity, and refining technical phrasing in the Introduction and Literature Review sections**.