AI Personal Trainer

Michael Cheung

1 Abstract

Regular physical exercise is important for health and wellness, but performing exercise with incorrect form can cause injuries. This paper presents a machine learning method to classify the quality of squats. Using an iPhone to record exercise, the videos are processed using Google's Mediapipe, which extracts x, y, z coordinates of body joints, which are then converted into joint angles, in particular, the knees and hips. A multilayer perceptron is trained to classify squat quality, achieving an accuracy of 92%. The results demonstrate the model's potential for real time feedback in personal fitness applications, providing a solution for improving exercise safety and effectiveness.

2 Introduction

According to the National Security council [1], an estimated 482,886 injuries, related to exercise and equipment, were treated in emergency department visits in 2023. Providing real time feedback on exercise with machine learning could significantly reduce these injury rates.

Correct exercise form is important because it reduces injuries and improves the effectiveness of workouts. Squats are a fundamental exercise for building lower body strength and muscle. But squatting incorrectly can lead to injuries, such as in the knee, hip, and lower back. Squats were selected due to their popularity and importance in strength training.

This project aims to develop a machine learning model that can classify squat quality from video recordings. The input to our algorithm is a video recorded by an iPhone, and the output is a binary classification indicating whether the squat is performed with good or bad form. We use Google's MediaPipe to extract the coordinates of key body points from the video. Next, we use an angle formula to calculate joint angles based on key body points. The angles are used as features for training the Multilayer Perceptron model. Our approach automates the assessment of squat quality and can be integrated into smartphones.

The key technical insight is the combination of pose estimation with machine learning to evaluate exercise form. Unlike existing solutions that can count repetitions, the model provides feedback on the quality of the exercise.

3 Related work

Pose estimation is a significant area of research in computer vision.

Toshev et al. [2] were the first to use deep neural networks for pose estimation in 2013. They did pose estimation with a cascade of deep neural network based regression. Since then, more accurate and sophisticated models have been developed. Newell et al. [3] used a stacked hourglass neural network. Shotton et al. [4] did pose estimation without the use of RGB cameras, instead using the Microsoft Kinect. They used depth maps from depth cameras to predict 3D body joint positions. There was also research done on pose estimation for multiple people in a single image. Cao et al. [5] developed OpenPose, a multi-person Pose Estimation using Part Affinity Fields.

This paper uses the pose estimation model from Google's MediaPipe API [6], which is an implementation of BlazePose [7], which provides on-device real-time body pose tracking. BlazePose is a lightweight convolutional neural network architecture specifically designed for mobile devices.

We explore the application of pose estimation in sports and fitness. Related work by Apple [8] offers APIs that classify an exercise type and counts repetitions. But, currently, there are no good solutions that can assess the quality of exercise.

4 Dataset and Features

Our dataset consists of 100 videos of the author performing squats, recorded with the iPhone's rear camera facing the subject. The dataset was split into 70% for training and 30% for testing. The videos were processed using MediaPipe, which outputs x, y, z coordinates and confidence scores for 33 body joints, such as the left knee, right knee, left hip, and right hip.

Next, feature engineering was done. Coordinates were not directly used due to the variations in squat done and camera angles. Instead, joint angles were calculated, particularly of the knees and hips because these joints are relevant in determining squat quality.

The knee and hip angles can show the quality of the exercise form. In good squats, these angles display a smooth parabola with minimal noise. In bad squats, these angles display a more irregular pattern. This shows what the machine learning model is learning.

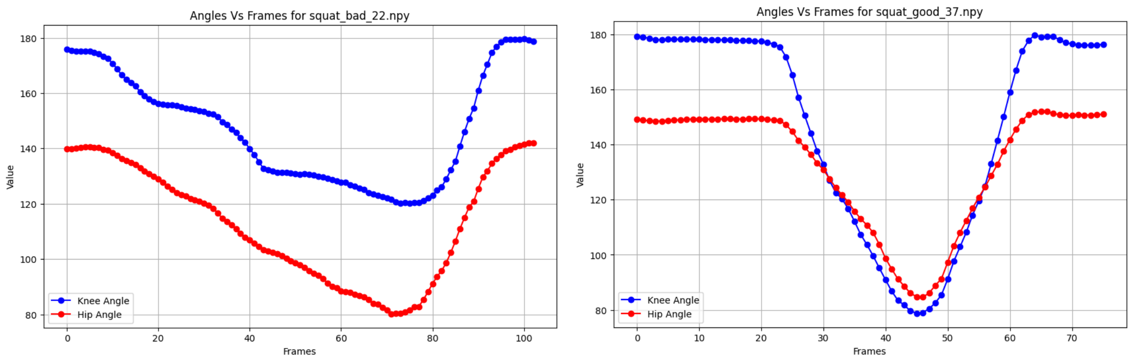

Figure 1. The first graph is a squat done with bad form and the second graph is one for good form. The graphs show how the knee and hip angles change during a squat.

Figure 2. Pose estimation is applied to a person in a seated position, which detects body landmarks.

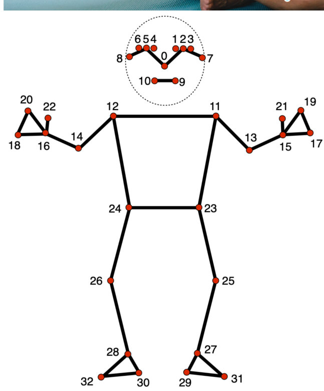

Figure 3. Pose estimation tracks 33 body landmark locations, representing the approximate location of the body parts.

5 Methods

We experimented with several machine learning models, such as Logistic Regression, Support Vector Machines, and Decision Trees. A Multilayer Perceptron was selected due to its performance.

The Multilayer Perceptron was implemented using the scikit-learn library [9] and consisted of a single hidden layer with 100 neurons. The activation function used in the hidden layer was the Rectified Linear Unit, to handle non-linear data effectively. The Adam optimizer was chosen for training. We performed hyperparameter tuning to identify the optimal learning rate, regularization, and number of neurons. The final model used a learning rate of 0.001, L2 regularization with a penalty parameter (α) of 0.0001, and a maximum of 200 training iterations.

The equation for the forward pass of the Multilayer Perceptron is described as follows:

$$\hat{y} = f(W_2 \cdot f(W_1 \cdot X + b_1) + b_2)$$

Where X is the input feature vector representing the joint angles of the knees and hips, W_1 and W_2 are the weight matrices for the hidden and output layers, b_1 and b_2 are the bias terms for the hidden and output layers, f is the ReLu activation function, and $\hat{y}$ is the predicted output for squat form quality.

6 Experiments/Results/Discussion

We note in the classification report that the model performs well. Our machine learning model achieves an accuracy of 90% indicating a good performance in classifying squat quality.

In the confusion matrix, we see that the model accurately classifies the correct class labels, but occasionally misclassifies instances of bad form as good form.

During training, we see that the loss curve reaches convergence after approximately 150 iterations. This indicates effective learning without overfitting.

	precision	recall	f1-score	support
0.0	0.77	1.00	0.87	10
1.0	1.00	0.85	0.92	20
accuracy			0.90	30
macro avg	0.88	0.93	0.89	30
weighted avg	0.92	0.90	0.90	30

Accuracy: 0.9

Figure 4. Classification report.

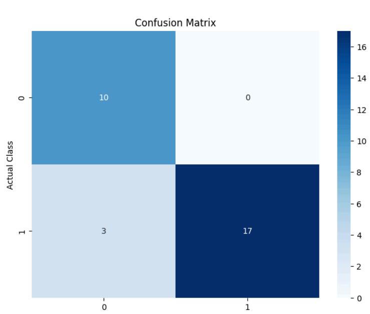

Figure 5. Confusion matrix.

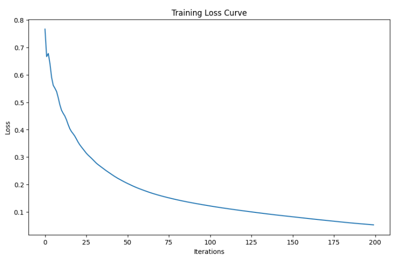

Figure 6. Loss curve.

7 Conclusion/Future Work

This project used machine learning to classify squat quality. The Multilayer Perceptron model achieved good performance metrics, suggesting its potential for real world applications in personal fitness.

There are many opportunities for future work. For example, expanding the dataset to include a more diverse range of participants with different body types and squat techniques could further improve the model's accuracy and generalizability. Additionally, integrating this model into a smartphone app that provides real-time feedback during workouts would help make personal fitness accessible to a greater number of people.

Further research could explore the use of more advanced models, such as Convolutional Neural Networks or Transformers, to capture both spatial and temporal dynamics of joint movements. Moreover, incorporating multimodal data, such as depth, could enhance the model's robustness in varying lighting conditions.

This project contributes to the field of AI fitness technology, offering a scalable and effective solution for improving exercise and reducing injuries.

8 References

1. “Sports and Recreational Injuries.” *Injury Facts*, 12 July 2024, injuryfacts.nsc.org/home-and-community/safety-topics/sports-and-recreational-injuries/#:~:text=For%20example%2C%20with%20regard%20to,than%20females%20(276%2C377%20males%20vs.
2. Toshev, Alexander, and Christian Szegedy. "Deeppose: Human pose estimation via deep neural networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.
3. Newell, A., K. Yang, and J. Deng. "Stacked hourglass networks for human pose estimation. arXiv." arXiv preprint arXiv:1603.06937 (2016).

4. Shotton, Jamie, et al. "Real-time human pose recognition in parts from single depth images." CVPR 2011. Ieee, 2011.
5. Cao, Zhe, et al. "Realtime multi-person 2d pose estimation using part affinity fields." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
6. “Pose Landmark Detection Guide | Google Ai Edge | Google AI for Developers.” Google, Google, ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker. Accessed 10 Sept. 2024.
7. Bazarevsky, V. "BlazePose: On-device Real-time Body Pose tracking." arXiv preprint arXiv:2006.10204 (2020).
8. “Detecting Human Body Poses in Images.” Apple Developer Documentation, developer.apple.com/documentation/Vision/detecting-human-body-poses-in-images. Accessed 10 Sept. 2024.
9. Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.