computers

MDPI

Article

Non-Standard Squat Posture Detection Method Using Human Skeleton

Leiyue Yao, Zhiqiang Dai and Keyun Xiong *

School of Intelligent Medicine and Information Engineering, Jiangxi University of Chinese Medicine, Nanchang 330100, China; daizhiqiang@jxutcm.edu.cn (Z.D.)
* Correspondence: 20032023@jxutcm.edu.cn

Abstract

Squats are essential for assessing lower limb strength. However, performing them incorrectly without professional guidance often leads to sports injuries. Currently, most detection methods rely heavily on deep neural networks and massive datasets. This approach brings several downsides. It involves high data labeling costs and heavy computing demands. It is also difficult to achieve low-latency feedback on mobile devices. Furthermore, these models often lack robustness when dealing with individual body differences. To tackle these issues, we propose a new real-time squat detection method. Our approach is built on prior rules and statistical models. Here is how it works. First, we use MediaPipe to track the body's skeleton joints in real-time from video feeds, calculating the hip and knee angles frame by frame. Next, we build a hip-knee coordination model using linear regression. This step helps us measure how these joints move together dynamically. Finally, we verify the squat depth using a geometry-based tolerance mechanism. This feature accounts for measurement noise and natural body variations, allowing us to accurately judge if the overall posture is standard. We tested our approach on three different squat styles. The results show that our method catches improper forms quickly and efficiently in real time, achieving an accuracy of 90%.

Keywords: squat recognition; posture detection; human skeleton analysis; linear regression; lightweight real-time detection model

1. Introduction

As the "king of strength training", the squat is regarded as a gold-standard exercise for evaluating lower extremity strength, core stability, and full-body coordination, and has been widely adopted in rehabilitation training and competitive sports. However, the squat involves complex coordinated movements of multiple joints, including the hip, knee, and ankle. Long-term performance of non-standard postures (such as knee valgus, insufficient squat depth, and discoordination between hip and knee motion) can easily lead to sports injuries, including meniscus wear [1], patellofemoral pain syndrome [2], and lumbar disc herniation [3]. Constrained by the high cost of professional personal trainers, ordinary fitness enthusiasts often rely on subjective perception to evaluate their movements, making it difficult to identify subtle postural deviations. Therefore, the automatic detection and real-time correction of squat postures using computer vision technology is of great practical significance for reducing exercise risks and improving training effectiveness.

With the rapid advancement of information technology, motion recognition methods have primarily evolved into two technical pathways: contact-based and contact-free [4,5].

Check for updates

Academic Editor: Paolo Bellavista

Received: 24 March 2026
Revised: 22 April 2026
Accepted: 29 April 2026
Published: 5 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Early research heavily relied on wearable sensors; however, these sensors require attachment to specific body parts, which causes discomfort during movement, impairs athletic performance, and are susceptible to damage from sweat, coupled with cumbersome data collection procedures. Consequently, this direct human-contact approach faces significant barriers to large-scale popularization in the context of general daily fitness training. Presently, contact-free detection based on computer vision has emerged as the dominant paradigm. According to the source of the input signal, existing vision-based detection methods are mainly categorized into two types: RGB-D (RGB-Depth) cameras and monocular RGB cameras.

Among depth camera-based solutions, the Kinect has been the device of choice in the field of motion capture. For instance, Kim et al. [6] employed the Kinect to recognize asymmetric gait and extract skeletal keypoint features; Sejera et al. [7] combined support vector machines to evaluate body angle variations during resistance training; and Wu et al. [8] used a Kinect motion analysis system to quantify bradykinesia in Parkinson's patients. However, when applied to complex compound movements such as squats, this approach still faces technical bottlenecks. First, severe self-occlusion often occurs when the body descends to the lowest squat position, leading to nonlinear drift in the skeletal model output by the depth sensor and deviation from the actual human contour. Second, as a fixed hardware device, the Kinect requires preinstallation and suffers from inherent limitations such as restricted recognition range and difficulty in mobile deployment, which cannot meet the demand for popularized and convenient squat posture detection.

For monocular RGB camera-based schemes, researchers have widely adopted deep learning or rule-based recognition methods. In gesture recognition, studies have continuously optimized recognition accuracy using deep learning models such as improved convolutional neural networks (CNNs) [9–11]. In sitting posture detection, vision-based deep learning models enable real-time posture assessment via cameras [12–14]. In fall detection, lightweight visual convolutional networks have effectively improved real-time recognition efficiency in surveillance footage [15–17]. Furthermore, in sports action recognition, methods based on 2D skeleton extraction and image feature matching have further expanded the application boundaries of mobile-oriented visual recognition [18,19]. Nevertheless, existing monocular RGB approaches still exhibit two notable limitations for the specific application of mobile squat detection: 1. Computational redundancy of deep learning: Although CNN-based visual techniques (e.g., VGG-16, ResNet) deliver high recognition accuracy, they rely heavily on large-scale annotated datasets and involve an extremely large number of model parameters. Such methods demand intensive computing resources and thus struggle to achieve low-latency real-time feedback on mobile phones. 2. Rigidity of traditional rule-based methods: Early rule-based approaches typically depend on prior knowledge by setting fixed angle thresholds or artificially dividing continuous movements into discrete phases such as "descent" and "ascent". This ignores individual differences in body proportions across populations. Meanwhile, staged judgment disrupts the temporal correlation among joints throughout the entire motion, making it unable to effectively capture complex coordination errors.

To overcome the aforementioned hardware barriers and algorithmic limitations, this paper proposes a lightweight squat detection method that only requires the color camera of a smartphone, while possessing both the positioning advantages of depth cameras and high computational efficiency. For low-level data acquisition, this work adopts Google's open-source MediaPipe framework [20], whose lightweight network architecture enables high-precision real-time skeleton extraction with monocular RGB cameras, delivering excellent spatial positioning accuracy and good compatibility with mobile devices [21,22]. For high-level decision logic, instead of using fixed static thresholds or large-scale neural

networks, this study returns to the essence of biomechanics and no longer divides squats into mechanical phases, but treats them as a continuous dynamic process, with the main innovations as follows:

1. Dynamic modeling of hip-knee coordination based on linear regression: Aiming at the limitation that conventional methods fail to adequately account for inter-joint synergism, this paper draws on the classical angle-angle diagram analysis concept from biomechanics, and refers to the work of Fuglsang et al. [23], who applied linear regression to evaluate limb motion characteristics. We innovatively introduce a linear regression model to quantify the dynamic correlation between hip and knee joint angles throughout the entire squat movement. By extracting the slope feature of the hip-knee angular change rate, the proposed algorithm achieves "de-absolutization" of the evaluation criterion. Specifically, the model no longer relies on the absolute height or limb length of individual subjects, but focuses instead on the relative rate of joint motion. This design enables the model to accurately identify synergistic movement errors such as good-morning squats (excessive hip-dominant motion) and knee-dominant squats, and significantly improves the generalizability of the algorithm across populations with diverse body types.
2. Robust depth verification based on geometric tolerance: In accordance with Schoenfeld's [24] biomechanical description of squat movements, the parallel squat (defined as the position where the thigh is parallel to the ground) is widely recognized as a standard squat depth. Nevertheless, Myer et al. [25] point out that a uniform squat depth standard is difficult to generalize for practical detection across individuals. To address this theoretical challenge as well as the joint coordinate jitter caused by mobile device cameras, this paper designs a depth judgment mechanism based on relative position. By calculating the vertical heights of the hip and knee joints and introducing a geometric tolerance, this mechanism allows for a reasonable margin of measurement error and effectively offsets data noise introduced by camera jitter. As a result, the proposed method not only adheres to established biomechanical principles but also greatly improves the system's robustness in real-world application scenarios.
3. Lightweight white-box detection architecture for mobile devices: Current mainstream deep learning-based human action recognition models, such as LSTM, STGCN, and VGG16, generally suffer from drawbacks including excessive parameter volume and substantial computational redundancy. Against this background, this paper innovatively constructs a hierarchical architecture of bottom-layer lightweight keypoint extraction + top-layer geometric statistics. Instead of adopting a black-box end-to-end deep neural network paradigm, the proposed structure does not require massive computational overhead. While maintaining an overall recognition accuracy of 91.67%, our method achieves an ultra-high inference speed of 98,000 FPS. This greatly lowers the deployment barrier for the algorithm on resource-constrained end devices such as ordinary mobile terminals, and provides a feasible, lightweight paradigm for low-latency, real-time movement correction in popular mass fitness scenarios.

The rest of the paper is structured as follows. Section 1 presents the research background and significance, and elaborates the proposed method for abnormal squat posture detection. Section 2 describes the experimental design and analyzes the experimental results. Section 3 concludes the whole work and prospects the future research directions.

Figure 1 illustrates the complete process, from recording a squat video to evaluating whether it is a standard movement. This process consists of three main steps:

- Step 1: A continuous sequence of squats is recorded using a mobile device and then segmented into individual squat clips, ensuring each clip contains one complete descent-and-ascent movement.

- Step 2: The MediaPipe framework is used to extract 33 human skeletal keypoint coordinates frame by frame for each action, and the left and right hip joint and knee joint angles are calculated for each frame of every sample (see Section 1.2 for details).
- Step 3: 80% of the samples in the standard squat category are selected to fit the slope, intercept, and threefold standard deviation of the left and right curves. The final fitting curve is obtained by averaging these values (see Section 1.3.1 for details). Combined with the designed squat depth judgment rule (see Section 1.3.2 for details), the entire model is constructed. A test sample is classified as a standard movement only if it passes both judgments.

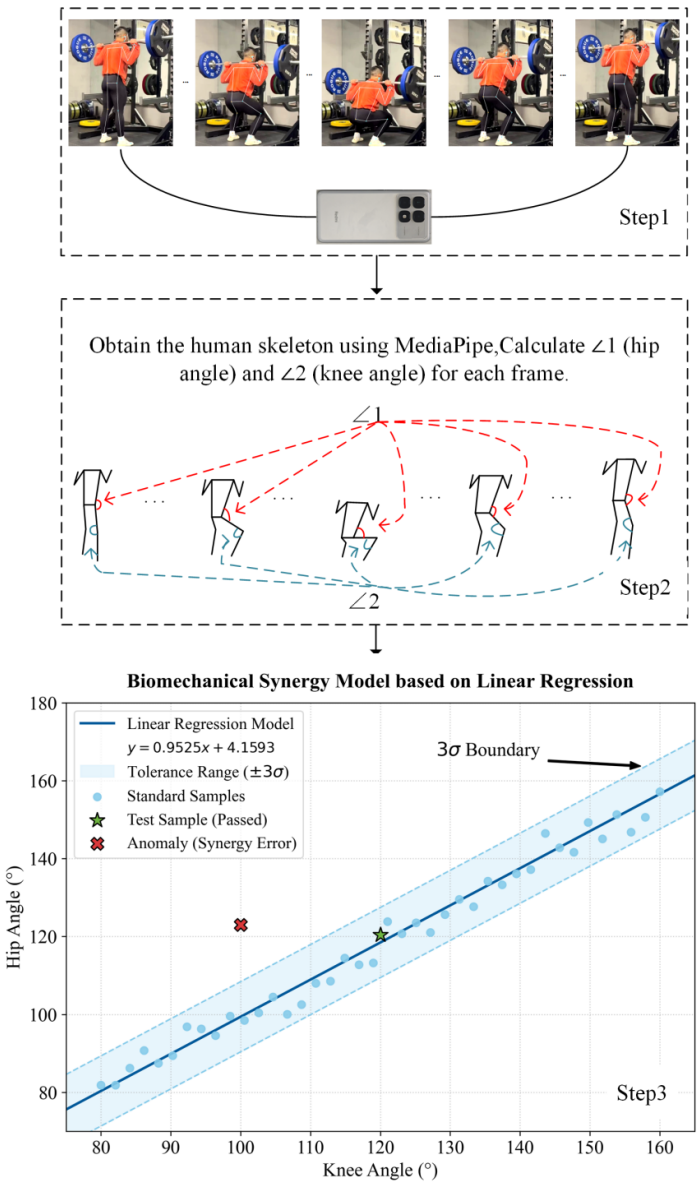


Figure 1. Overall Flowchart.

1.1. MediaPipe

This study adopts Google's open-source MediaPipe framework as the underlying visual processing core. Compared with conventional alternatives such as OpenPose and PoseNet, MediaPipe features a lightweight network architecture that is specifically optimized for real-time inference on mobile and edge devices. We selected this framework based on three primary considerations. First, MediaPipe's highly efficient dataflow process-

ing meets the low-latency requirements of frame-by-frame motion analysis in the proposed system. Second, the framework exhibits strong robustness for skeleton extraction under varying illumination and occlusion conditions, making it well-suited for indoor gym environments. Third, MediaPipe's cross-platform compatibility enables direct deployment on mobile devices without relying on specialized acquisition hardware such as Kinect, which significantly lowers the barrier to practical usage for end users.

1.2. Key Joint Acquisition and Angle Calculation

The primary technical challenge addressed in this study is to capture human skeleton sequences in real time and with high accuracy for each frame. Although depth sensors such as Kinect are widely used, the estimated joint positions suffer from noticeable drift under non-ideal conditions, including single-view capture and excessively short shooting distances. To mitigate this issue, the industry commonly adopts multi-camera cooperative calibration to improve positioning accuracy. However, such an approach not only incurs high hardware deployment costs but also restricts the user's free movement space due to the complex calibration procedure, making it difficult to meet the convenience requirements of general users. In contrast, the machine learning-based MediaPipe framework demonstrates superior cross-scene robustness. With its high-performance inference engine, this framework can still accurately regress 33 structured human skeleton joints even under environments of dim indoor light, complex backgrounds, or limb occlusion, as illustrated in Figure 2.

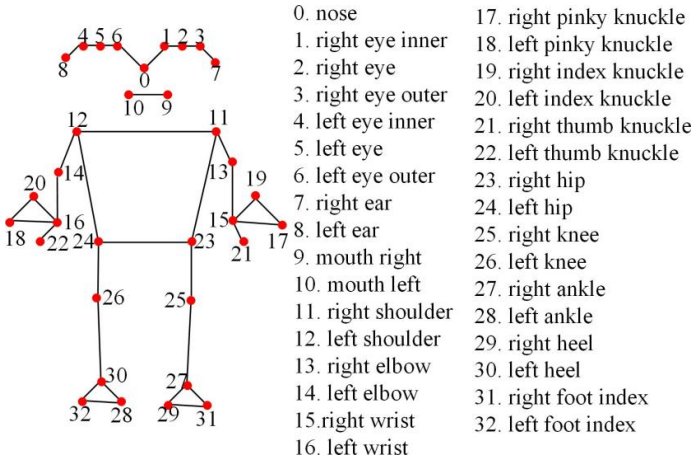


Figure 2. Human Skeleton Joint Points Obtained by MediaPipe.

The two squat features constructed in this study can be calculated based on the law of cosines using 8 of the 33 joint points: right shoulder, left shoulder, right hip, left hip, right knee, left knee, right ankle, and left ankle.

As shown in Figure 3, for the hip joint angle, this study utilizes the right shoulder, left shoulder, right hip, left hip, right knee, and left knee to construct and calculate it. The calculation formulas are given in Equations (1) and (2). Here, $H(X_h, Y_h)$, $S(X_s, Y_s)$, and $K(X_k, Y_k)$ represent the coordinate information of the hip joint, shoulder joint, and knee joint in the current frame, respectively. Then, the hip joint angle can be calculated through Equation (3). The knee joint angle can also be obtained using the same method.

$$\vec{HS} = (X_s - X_h,\ Y_s - X_h) \tag{1}$$

$$\vec{HK} = (X_k - X_h,\ Y_k - X_h) \tag{2}$$

$$\cos(a) = \cos(\vec{HS}, \vec{HK}) = \frac{\vec{HS} \cdot \vec{HK}}{|\vec{HS}| \cdot |\vec{HK}|} \quad (3)$$



Figure 3. Diagram of Two Characteristic Angles.

1.3. Squat Standardization Determination Algorithm Based on Geometric Statistical Model

Traditional rule-based methods are sensitive to individual anatomical differences and fail to capture dynamic inter-joint coordination characteristics. To address these limitations, this paper proposes a dual evaluation mechanism comprising dynamic hip-knee coordination modeling and robust depth validation. The proposed algorithm regards the squat as a continuous spatiotemporal dynamic process, rather than evaluating it as a set of discrete static frames.

1.3.1. Hip-Knee Synergy Determination Based on Linear Regression

When squatting down, the linkage rule of the hip joint angle and knee joint angle is a big biomechanical feature of the depth action. Traditional methods often just judge a single angle threshold for static evaluation, ignoring the relative movement rates between joints. Consequently, it is hard to find coordination errors, like facing a "good morning squat" (that is, the knee joint is locked and the hip joint is over-flexed). Inspired by the "angle-angle diagram" analysis method in biomechanics [23], this paper proposes to use a linear regression model to quantify the dynamic mapping relationship of the hip and knee joints in the standard squat process. Suppose the knee joint angle at time t is $\theta_{knee}(t)$ and the hip joint angle is $\theta_{hip}(t)$, we build the linear regression model of Equation (4).

$$\theta_{hip}(t) = k \cdot \theta_{knee}(t) + b \quad (4)$$

Among them, k is the slope and b is the intercept. Through the standard dataset built by this paper, we got the optimal parameters as $k = 0.9525$ and $b = 4.1593$ (k, b, and 3σ are all obtained by averaging the values from the left and right sides). The slope k near 1 means that in a standard squat, the hip joint and knee joint change at almost the same rate. To adapt to different experimenters' body shapes and reduce the error brought by camera shake when shooting, this paper uses a statistical fault tolerance threshold. Calculate the residual between the actual observation angle and the theoretical prediction value. If the residual is in the 3σ statistical range, then judge this frame's coordination as qualified. The judgment condition is like Equation (5):

$$|\theta_{hip}^{\text{actual}}(t) - \theta_{hip}^{\text{predicted}}(t)| \leq 3\sigma \quad (5)$$

where the tolerance value is set to 15°. This value derives from the upper bound of the 3σ residual distribution of standard samples. Consequently, this model effectively

transforms action discrimination. It shifts the focus from an "absolute angle" to a "relative synergistic rule." This significantly enhances the algorithm's generalization capability across different populations.

1.3.2. Squat Depth Validation Based on Geometric Tolerance

Squat depth is another key index to evaluate squat standardness. According to Schoenfeld's [24] definition of "parallel squat", the thigh surface should be at least parallel to the ground. At this time, the hip joint and knee joint are on the same horizontal plane.

But in actual detection, everyone's mobile phone placement height and angle are different. Strictly requiring the "hip lower than knee" hard threshold will lead to a higher false negative rate (that is, misjudging qualified actions as unqualified).

To resolve this issue, this paper introduces the concept of soft constraints from fuzzy logic. We designed a depth evaluation mechanism based on geometric tolerance. The algorithm first identifies the frame representing the lowest point of the squat. Let the vertical coordinate of the hip joint in this frame be Y_{hip}, and the vertical coordinate of the knee joint be Y_{knee}. The decision condition is defined in Equation (6):

$$Y_{hip} \geq Y_{knee} - \delta \tag{6}$$

In this equation, δ represents a slack variable. We define this variable as an adaptive proportional value within the pixel coordinate system. This inequality means that the system considers the depth acceptable as long as the hip is not visually significantly higher than the knee. Specifically, it allows the hip to be parallel to or slightly above the knee, provided it remains within the defined tolerance range. This mechanism ensures that the movement strictly adheres to biomechanical standards. Simultaneously, it effectively offsets measurement errors caused by sensor noise. Consequently, it significantly enhances the overall robustness of the system.

2. Experiments and Results

2.1. Experimental Environment and Dataset

This paper's experiment is based on the following software and hardware. We use a desktop computer with an AMD Ryzen 7 3700X CPU (Advanced Micro Devices, Inc., Santa Clara, CA, USA), an NVIDIA GeForce GTX 1080 Ti 11GB graphics card (NVIDIA Corporation, Santa Clara, CA, USA), 64 GB of memory, and a Windows 10 operating system (Microsoft Corporation, Redmond, WA, USA). The experimental platform is PyCharm 2024 (JetBrains s.r.o., Prague, Czech Republic), and the development environment is Python 3.10 (Python Software Foundation, Wilmington, DE, USA). The shooting equipment is a Redmi K70 smartphone (Xiaomi Corporation, Beijing, China).

The experimental dataset constructed in this study covers three categories of squat movement postures, with a total sample size of 344 cases. Among them, there are 299 cases of standard squat postures as positive samples and 45 cases of two typical incorrect postures as negative samples. A total of 35 subjects were recruited for the study, including 31 males and 4 females, with an age distribution of 20–40 years. The physical parameters of the subjects are as follows: height ranges from 1.50 to 1.85 m, and body mass ranges from 50 to 80 kg. Data collection was performed using mobile devices. The height of the devices from the ground was controlled at 0.8–1.5 m, and the distance between the subjects and the recording devices was 1–3 m. Furthermore, to ensure the generalizability of the algorithm in real-world mobile applications, data collection was not limited to a strictly fixed single angle. All squat videos of the 35 subjects were recorded by the participants themselves, and no uniform shooting angle was imposed on them. The number of frames used to fit

the curve is all frames from the samples, covering the entire squat-stand cycle, with a total of 18,140 frames.

2.2. Data Splitting Strategy

To objectively evaluate the generalization ability of the model and prevent data leakage caused by repeated subject identities, this study adopted a "subject-based identity grouping" strategy. The specific splitting logic is as follows:

1. Splitting of positive samples (standard squats): All standard squat samples were grouped by subject ID. Data from 80% of the subjects were allocated to the training set. This set was used to fit the linear regression model parameters (k, b). It also calculated the statistical tolerance threshold (σ). Data from the remaining 20% of the subjects formed the validation set. This set evaluated the model's recognition rate for unseen standard movements.
2. Splitting of negative samples: This study adopts an anomaly detection approach based on a normal distribution. Consequently, the model only needs to learn "what is correct." Therefore, incorrect movement samples (shallow and good morning squats) were excluded from the training phase. Instead, they were entirely assigned to the test set. This verified the model's interception capability against different types of incorrect movements.

2.3. Experimental Results and Analysis

The proposed algorithm was utilized to verify the three squat postures in the test set, and the experimental results are shown in Table 1.

Table 1. Results of Squats with Three Different Postures.

Category	Sample Size	Hard Constraint	Soft Constraint	Final Pass Rate
Standard Squat	72	98.61%	100%	98.61%
Shallow Squat	38	23.68%	100%	23.68%
Good Morning Squat	10	0%	0%	0%

By comparatively analyzing the judgment performance of the "hard constraint (depth)" and the "soft constraint (synergy)," the following conclusions can be drawn:

1. Performance on standard squats: On the validation set consisting of 72 standard squat samples, the system achieves a recognition accuracy of 98.61%, and the pass rate under the soft constraint (coordination judgment) reaches 100%. This demonstrates that the linear regression model based on the 3σ statistical interval can adapt to subjects with diverse body types and effectively compensate for joint coordinate jitter detected by MediaPipe.
2. Performance on shallow squats: For the 38 shallow squat samples, the pass rate under the hard constraint is merely 23.68%, which effectively rejects most movements that fail to meet the required depth standard. Meanwhile, the pass rate of these samples under the soft constraint reaches 100%. This result carries important biomechanical implications, demonstrating that the proposed algorithm can independently evaluate motion trajectory and motion range. In other words, the system can identify that the subject's hip-knee coordination mechanism is correct, while the squat depth fails to reach the parallel standard.
3. Performance on "good morning" squats: In the test for "good morning squat" movements, the algorithm successfully rejected 100% of such samples. In this type of movement, the hip extension rate is significantly faster than the knee extension

rate, which severely violates the linear linkage rule of $k = 0.9525$. Consequently, the residual far exceeds the 3σ tolerance threshold. This indicates that the linear regression–based dynamic modeling is highly sensitive for detecting such high-risk compensatory movements.

Figure 4 illustrates the four performance metrics of our model.

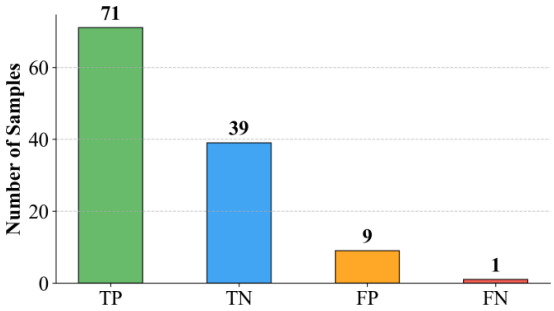


Figure 4. Data Result Graph.

2.4. Comparison with Mainstream Methods

To further verify the effectiveness and lightweight advantages of the proposed algorithm, we conducted a performance comparison. We compared our approach against current mainstream deep learning models for action recognition. These models include LSTM, ST-GCN, VGG-16, ResNet-50, and 1D-CNN. The specific experimental results are presented in Table 2.

Table 2. Comparison of Deep Learning Methods and the Proposed Method.

Method	TP	FP	FN	TN	Precision	Recall	F1-Score	Accuracy
LSTM	54	13	0	0	80.60%	100.00%	0.8926	80.59%
ST-GCN	48	1	6	12	97.96%	88.89%	0.9320	89.55%
VGG16	48	3	6	10	94.12%	88.89%	0.9143	86.56%
RESNET-50	43	0	11	13	100.00%	79.63%	0.8866	83.58%
1D-CNN	54	13	0	0	80.60%	100.00%	0.8926	80.59%
Proposed	71	9	1	39	88.75%	98.61%	0.9342	91.67%

First, regarding recognition accuracy, the proposed method achieves 91.67%, which comprehensively outperforms conventional models, including LSTM (80.59%), VGG-16 (86.56%), and 1D-CNN (80.59%). It even slightly surpasses ST-GCN (89.55%), a highly acclaimed model in the field of graph convolutional networks. Since the MediaPipe framework regresses internal joint coordinates from external body contours, loose clothing severely distorts the perceived human silhouette. This causes the estimated hip joint coordinates to be artificially lowered in the 2D image, thereby incorrectly satisfying the geometric depth criterion even when the actual anatomical posture is inadequate.

Second, regarding real-time processing capability (Frames Per Second, FPS), our method demonstrates a significant advantage, as shown in Table 3. Vision technologies based on CNNs offer high recognition accuracy. However, they contain massive parameter counts (such as VGG-16). This creates significant computational redundancy. Consequently, they impose extreme hardware requirements on mobile devices and struggle to meet the demand for real-time feedback. As seen in Table 3, the processing speeds of VGG-16 and ResNet-50 are only 30 FPS and 73 FPS, respectively. Even the fastest 1D-CNN model achieves only 3450 FPS. In contrast, our proposed method reaches 98,000 FPS.

Table 3. Comparison of Speed.

Method	FPS
LSTM	177
ST-GCN	955
VGG16	30
RESNET-50	73
1D-CNN	3450
Proposed Method	98,000

Compared to other methods, this paper's method performance shows a huge improvement. The basic reason is the simple algorithm structure. In the bottom layer, we adopt the lightweight MediaPipe framework for real-time human skeleton extraction. In the top layer, the system executes low-computing-cost linear regression and geometric coordinate comparison. This design not only keeps a high accuracy rate but also greatly reduces the computing power redundancy brought by deep learning. This makes our algorithm have the deployment potential to realize low delay on mobile smartphones.

2.5. Hard Case Analysis

The proposed algorithm achieves a recognition accuracy of 91.67% on the overall test set. However, due to the inherent physical limitations of monocular 2D vision, a small number of misclassified samples still exist. We performed frame-by-frame playback and visual analysis on the hard misclassified samples (including false negatives and false positives). Further investigation reveals that these "hard examples" do not fail randomly, but expose the system's sensitivity to complex real-world factors—specifically, sensitivity to critical boundaries and occlusion caused by clothing. First, for critical actions that lie exactly on the qualification boundary, the system is highly vulnerable to hardware-induced measurement noise and self-occlusion. At the lowest squat position, these factors cause subtle coordinate shifts, which further lead to misjudgment and incorrectly reject borderline standard movements. Second, the system's dependence on external visual contours makes it susceptible to human silhouette distortion. For instance, detailed analysis shows that all 9 false positive samples came from one test subject wearing excessively loose lower-body clothing. Since the MediaPipe framework regresses internal joint coordinates from external contours, loose clothing distorts the human silhouette, causing the estimated hip joint coordinates in the 2D image to be lower than the actual position and erroneously satisfy the geometric depth criterion. In summary, these hard examples do not indicate a failure of the biomechanical rule-based logic proposed in this paper, but highlight the physical limitations of current mobile-based pose estimation methods. To address these issues, future versions of the system will focus on introducing temporal smoothing filters to reduce coordinate jitter caused by hardware. In addition, as mentioned in the Section 3, higher-precision pose estimation frameworks such as OpenPose will be evaluated and compared to further improve recognition accuracy.

3. Conclusions and Future Work

Aiming at the problem that squats easily cause sports injuries in daily training because of non-standard postures (like hip and knee coordination disorder, not enough squat depth), this paper proposes a lightweight squat detection method based on MediaPipe and a geometric statistical model. Through MediaPipe, we capture human skeleton data in real time and extract hip joint and knee joint angles. Combining the linear regression

coordination modeling and geometric tolerance judgment mechanism, it realizes real-time and high-efficiency judgment of squat action standardness. The experiment results show that, under the premise of not needing massive training samples and expensive hardware equipment, this method realizes a 91.67% recognition accuracy rate and very high inference speed. It has good practicality and mobile terminal promotion value. This paper's dataset can be downloaded at: https://doi.org/10.5281/zenodo.19550070.

Meanwhile, the proposed method still has certain limitations. For instance, the dataset is captured via mobile devices, and human skeletal keypoints are detected using MediaPipe. Such hardware-based acquisition inevitably introduces a certain degree of measurement noise. In future work, we plan to adopt methods such as Kalman filtering or bicubic interpolation to correct the measurement errors that may occur in each frame. These issues are not addressed in the present paper.

For the rule-based method, a prerequisite application is the use of MediaPipe, which is lightweight and efficient, making it well-suited for mobile terminal requirements. However, the accuracy of the framework directly affects the extraction of joint coordinates. In the future, we will explore high-precision frameworks with stronger anti-distortion capabilities, such as OpenPose, compare their errors with those of MediaPipe, and verify the effect of high-precision information on reducing misjudgments. This study still has certain limitations. The time consumption of human skeleton keypoint extraction by MediaPipe is not included when calculating FPS. In future work, with the increasing maturity of human keypoint acquisition techniques, the overall efficiency of the proposed method can be further improved.

Author Contributions: Conceptualization, L.Y.; methodology, Z.D.; validation, Z.D.; writing—original draft preparation, Z.D.; writing—review and editing, L.Y.; supervision, L.Y.; funding acquisition, K.X. All authors have read and agreed to the published version of the manuscript.

Funding: National Natural Science Foundation of China: 62366023.

Data Availability Statement: The raw video recordings of the subjects are not publicly available due to privacy and ethical restrictions regarding identifiable human subjects. However, the de-identified processed data, including skeletal keypoint coordinates and joint angles, are openly available in Zenodo at https://doi.org/10.5281/zenodo.18859797.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kawada, K.; Yokoyama, Y.; Tamura, M.; Okazaki, Y.; Ozaki, T.; Furumatsu, T. Occupational motions such as kneeling and squatting are associated with theincreased development of medial meniscus posterior root tears. *J. Orthop. Sci.* **2025**, *30*, 106–114.
2. Xie, P.; István, B.; Liang, M. The relationship between patellofemoral pain syndrome and hipbiomechanics: A systematic review with meta-analysis. *Healthcare* **2023**, *11*, 99. [CrossRef] [PubMed]
3. García-Jaén, M.; Del Barrio-Ventura, A.; Sanchis-Soler, G.; Sebastiá-Amat, S. Exercise-based therapeutic interventions for lumbar disc herniation: A narrative review. *J. Phys. Educ. Sport* **2025**, *25*, 825–832.
4. Yin, Y.; Xie, L.; Jiang, Z.; Xiao, F.; Cao, J.; Lu, S. A Systematic Review of Human Activity Recognition Based on Mobile Devices: Overview, Progress and Trends. *IEEE Commun. Surv. Tutor.* **2024**, *26*, 890–929. [CrossRef]
5. Shin, J.; Hassan, N.; Miah, A.S.M.; Nishimura, S. A Comprehensive Methodological Survey of Human Activity Recognition Across Diverse Data Modalities. *Sensors* **2025**, *25*, 4028. [CrossRef]
6. Kim, J.W.; Choi, J.Y.; Ha, E.J.; Choi, J.H. Human Pose Estimation Using MediaPipe Pose and Optimization Method Based on a Humanoid Model. *Appl. Sci.* **2023**, *13*, 2700. [CrossRef]
7. Sejera, M.M.; Arzobal, E.A.B.; Soliongco, J.G.R. Real-time deadlift form analysis using support vector machine algorithm and Kinect V2 sensor. In Proceedings of the 5th Eurasia Conference on IoT, Communication and Engineering (ECICE), Yunlin, Taiwan, 27–29 October 2023; pp. 722–726.
8. Wu, Z.; Gu, H.; Hong, R.; Xing, Z.; Zhang, Z.; Peng, K.; He, Y.; Xie, L.; Zhang, J.; Gao, Y.; et al. Kinect-based objective evaluation of bradykinesia in patients with Parkinson's disease. *Digit. Health* **2023**, *9*, 20552076231176653. [CrossRef]

9. Shanmugam, S.; Narayanan, R.S. An accurate estimation of hand gestures using optimal modified convolutional neural network. *Expert Syst. Appl.* **2024**, *249*, 123351. [CrossRef]
10. Murad, K.; Alasadi, A.H.H. Advancements and Challenges in Hand Gesture Recognition: A Comprehensive Review. *Iraqi J. Electr. Electron. Eng.* **2024**, *20*, 154–164. [CrossRef]
11. Dhiman, R.; Luthra, P.; Singh, N.T. Different categories of feature extraction and machine learning classification models used for hand gesture recognition systems: A review. In Proceedings of the 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), Lonavla, India, 7–9 April 2023; pp. 1–7.
12. Lin, B.S.; Liu, K.J.; Tseng, W.H.; Ahmed, A.M.; Wang, H.C.; Lin, B.S. A Deep Learning-Based Chair System That Detects Sitting Posture. *IEEE J. Biomed. Health Inform.* **2024**, *28*, 482–493. [CrossRef]
13. Hu, H.; Shi, X.; Song, W.; Yang, Y.; Zhang, J. Research Progress of Intelligent Sitting Posture Monitoring Systems: A Survey. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 9527923. [CrossRef]
14. Jin, H.; He, X.; Wang, L.; Zhu, Y.; Jiang, W.; Zhou, X. SitPose: Real-Time Detection of Sitting Posture and Sedentary Behavior Using Ensemble Learning with Depth Sensor. *IEEE Sens. J.* **2024**, *25*, 12444–12454. [CrossRef]
15. Owusu, E.; Acquah, I.; Asare, M.A.; Yeboah, B.A. LiteFallNet: A lightweight deep learning model for efficient real-time fall detection. *Digit. Health* **2025**, *11*, 20552076251386698. [CrossRef] [PubMed]
16. Jiang, Z.; Al-Qaness, M.A.; Al-Alimi, D.; Ewees, A.A.; Abd Elaziz, M.; Dahou, A.; Helmi, A.M. Fall Detection Systems for Internet of Medical Things Based on Wearable Sensors: A Review. *IEEE Internet Things J.* **2024**, *11*, 34797–34810. [CrossRef]
17. Liu, J.; Li, X.; Huang, S.; Chao, R.; Cao, Z.; Wang, S.; Wang, A.; Liu, L. A review of wearable sensors based fall-related recognition systems. *Eng. Appl. Artif. Intell.* **2023**, *121*, 105993. [CrossRef]
18. Cao, M.; Wan, J.; Gu, X. CLEAR: Multimodal Human Activity Recognition via Contrastive Learning Based Feature Extraction Refinement. *Sensors* **2025**, *25*, 896. [CrossRef]
19. Ko, Y.M.; Nasridinov, A.; Park, S.H. Real-time AI posture correction for powerlifting exercises using YOLOv5 and MediaPipe. *IEEE Access* **2024**, *12*, 195830–195853. [CrossRef]
20. Lugaresi, C.; Tang, J.; Nash, H.; McClanahan, C.; Uboweja, E.; Hays, M.; Zhang, F.; Chang, C.L.; Yong, M.; Lee, J.; et al. MediaPipe: A framework for perceiving and processing reality. In Proceedings of the 3rd Workshop on Computer Vision for AR/VR at IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 17 June 2019.
21. Sirikongtham, P.; Nimkoompai, A. A new method for real-time fall detection based on MediaPipe pose estimation and LSTM. *Int. J. Adv. Comput. Sci. Appl.* **2025**, *16*, 107. [CrossRef]
22. Garg, S.; Saxena, A.; Gupta, R. Yoga pose classification: A CNN and MediaPipe inspired deep learning approach for real-world application. *J. Ambient Intell. Humaniz. Comput.* **2023**, *14*, 16551–16562. [CrossRef]
23. McKean, M.; Burkett, B.J. Does segment length influence the hip, knee, and ankle coordination during the squat movement? *J. Fit. Res.* **2012**, *1*, 23–30.
24. Schoenfeld, B.J. Squatting kinematics and kinetics and their application to exercise performance. *J. Strength Cond. Res.* **2010**, *24*, 3497–3506. [CrossRef]
25. Myer, G.D.; Kushner, A.M.; Brent, J.L.; Schoenfeld, B.J.; Hugentobler, J.; Lloyd, R.S.; Vermeil, A.; Chu, D.A.; Harbin, J.; McGill, S.M. The back squat: A proposed assessment of functional deficits and technical factors that limit performance. *Strength Cond. J.* **2014**, *36*, 4–27. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.