Journal of Intelligent System and Telecommunications

Journal homepage: https://journal.unesa.ac.id/index.php/jistel/index

Two-dimensional Human Pose Estimation using Key Points' Angular Detection for Basic Strength Training

Achmad Ivan Taruna Jaya[1], Pradini Puspitaningayu[2], Athaya Pradipa Adiwangsa[3], Nobuo Funabiki[4]*
[1,2,3]Dept. of Electrical Engineering, Faculty of Engineering, Universitas Negeri Surabaya, Surabaya, Indonesia
[4]Dept. of Electrical and Communication Engineering, Okayama University, Japan
*Correspondence E-mail: pradinip@unesa.ac.id

ARTICLE INFO

Article History:
Submitted/Received 23 December 2024
First Revised 31 December 2024
Accepted 31 December 2024
First Available Online 31 December 2024
Publication Date 31 December 2024

Keyword:
Human Pose Estimation, Sports Performance Analyis, MediaPipe, Deep Learning, Image Processing, Artificial Intelligence.

© 2024 Tim Pengembang Jurnal UNESA

ABSTRACT

Human pose estimation (HPE) has emerged as a crucial topic in computer vision, with applications ranging from sports training to injury prevention. This paper proposes a real-time 2D pose estimation system that leverages keypoint angle detection for basic strength exercises, such as squats, bicep curls, and deadlifts. The system integrates MediaPipe to detect joint positions and analyze them against optimal movement patterns determined by fitness guidelines. Primary data were collected using standard webcam recordings of exercises performed by expert trainers, enabling the system to establish joint angle thresholds and validate user movements. The system's performance was evaluated through user trials, where it successfully validated repetitions and provided real-time feedback. Results showed an average accuracy of 80% across all exercises, with sensitivity maintained at 100%. Pearson correlation analysis demonstrated strong validation performance, with a coefficient of 0.98. Factors affecting performance included discrepancies in body proportions and user familiarity with strength training techniques. Despite these challenges, the system effectively highlighted errors, promoting improved form and reduced injury risk. This study contributes to the development of accessible and efficient real-time HPE systems that can operate on standard hardware. It emphasizes the practical application of pose estimation in enhancing training outcomes, enabling independent users to improve their techniques while minimizing injury risks. Future work will expand the range of exercises and integrate automatic exercise detection to further improve the system's usability and versatility.

1. INTRODUCTION

Human pose estimation (HPE) has become a significant topic in the field of computer vision, with various methods developed to detect and analyze the configuration of human joints from images or videos. This technology initially relied on traditional vision-based approaches for pose tracking and reconstruction [1]. These studies established the groundwork for comprehending the essential processes in HPE, encompassing detection, tracking, and the computation of spatial relationships among body points. The development of machine learning technologies has brought significant advancements in HPE, enabling models to achieve higher accuracy by leveraging more complex data [2][3][4]. Frameworks like the Stacked Hourglass Network demonstrate how hierarchical approaches can enhance pose detection performance MPII Human Pose Dataset[5][6]. PoseAnalyser, Providing an overview of the architectures and algorithms used for 2D and 3D HPE, including methods such as CNN, OpenPose, and MediaPipe.[7].

A significant breakthrough in HPE occurred with the introduction of the DeepPose method, which employs a deep neural network-based approach to directly regress human pose coordinates with high accuracy [8][9][10]. This method was one of the first to employ direct regression using CNNs for HPE. Subsequently, Convolutional Pose Machines were developed, using a modular architecture to learn the spatial relationships between body points progressively, thereby achieving higher accuracy in pose estimation tasks [11]. Advanced models like HRNet further improved accuracy by maintaining high-resolution representations throughout the inference process [12].

The HPE libraries, such as OpenPose, PoseNet, MoveNet, and MediaPipe, have been compared in previous studies to evaluate their performance in detecting body key points from static images and videos. The results indicate that MoveNet demonstrates the best performance in accurately detecting human poses, particularly in the context of analyzing complex movements [13][14]. Modern frameworks such as MediaPipe [15] and OpenPose [16] have made significant contributions by enabling real-time pose estimation. MediaPipe provides an efficient pipeline for lightweight devices such as smartphones and tablets, while OpenPose supports multi-person pose estimation using the Part Affinity Fields (PAF) approach [17]. These two frameworks have become benchmarks for general-purpose pose estimation applications. Additionally, MediaPipe's flexibility allows it to support web-based and mobile applications [7][18].

In the context of sports, the application of HPE has become increasingly relevant for analyzing movements and improving athletic performance [19]. For instance, 3D HPE methods based on video leverage temporal convolutions to capture the dynamics of continuous movements [2]. This research offers a semi-supervised approach to enhance pose accuracy in physical activities, making it highly beneficial for sports involving complex motions. Other studies have also focused on the application of human pose tracking techniques in sports activities, considering variations in body posture that occur during dynamic movements [20]. Guidelines from the American Council on Exercise also make a significant contribution by establishing posture and technique standards for physical exercises like squats, bicep curls, and deadlifts. These guidelines are particularly relevant for evaluating the accuracy of movements performed during strength training [21].

The use of Node.js as a web-based server to execute pose evaluation functions highlights the importance of developing efficient and easily accessible systems for users. In the context of physical exercise, the integration of pose estimation and automated evaluation provides a

DOI: http://dx.doi.org/10.xxxxx/jistel.vXiX
e- ISSN xxxx-xxxx

practical solution that can be accessed in real-time, enabling users to receive immediate feedback while performing strength training or other physical activities [22][23].

Research has also explored the application of pose estimation in sports and exercise contexts, such as yoga posture recognition [24] and rehabilitation exercises [25]. Despite these advancements, there remains a noticeable gap in the implementation of real-time 2D pose estimation specifically designed for basic strength exercises, incorporating integrated joint angle analysis and feedback mechanisms.

This study aims to address this gap by developing a real-time system for 2D HPE using key point angle detection specifically for basic strength exercises. The objective of this research is to implement an efficient 2D pose estimation system using MediaPipe that accurately detects joint positions during strength exercises and analyzes them against optimal movement patterns determined by sports science guidelines. This system provides real-time feedback to users, helping them maintain proper form and reduce the risk of injury.

The research contributes to the development of a computationally efficient system suitable for real-time applications on standard hardware without requiring specialized equipment. This system enhances user experience by delivering quick and actionable feedback to improve exercise form and safety. It can help reduce the occurrence of exercise-related injuries and increase the effectiveness of strength training routines. Furthermore, fitness professionals can leverage this system as a tool for remote training and monitoring.

This paper is organized as follows: Section 2 explores relevant studies in literature, section 3 discusses the methods in the system, section 4 shows testing results, and Section 5 concludes this paper with future works.

2. RELATED WORKS

2.1 Human Pose Estimation

The authors in [1] discuss current improvements in vision-based motion capture technology in their comprehensive assessment. Their assessment systematically classifies different approaches, identifies ongoing obstacles, and outlines distinct applications in human motion analysis. This involves the identification, monitoring, and reconstruction of human movements from visual data, offering a thorough assessment of the existing landscape and proposing possible avenues for future research.

The study referenced in [2] explores the utilization of deep learning methodologies in HPE. The authors evaluate various neural network topologies, analyzing their efficiency and effectiveness. They emphasize critical problems, like accuracy and generalization across diverse datasets and contexts, providing insights into the trade-offs inherent in the design of these systems.

The authors in [4] introduce Recurrent Human Pose Estimation, which combines feed-forward and iterative modules. This architecture refines pose system predictions through recurrent learning, improving estimation accuracy by leveraging previous iterations to adjust predictions dynamically.

The authors in [5] made a significant addition to this discipline with the creation of the Stacked Hourglass Network. This hierarchical model is engineered to iteratively process multi-scale information, therefore substantially improving the precision of posture detection. The architecture's capacity to enhance predictions via iterative bottom-up and top-down processing illustrates an innovative method for managing spatial hierarchies in deep learning models.

DOI: http://dx.doi.org/10.xxxxx/jistel.vXiX
e- ISSN xxxx-xxxx

The authors in [6] develop DeepPose, the first approach to directly regress human pose coordinates using convolutional neural networks (CNNs). This study marks a significant breakthrough in deep learning for pose estimation.

The authors' groundbreaking research in [7], titled DeepPose, represented a substantial advancement in the application of deep neural networks for pose assessment. This approach use convolutional neural networks to directly regress pose coordinates from photos, establishing a core framework for future study in the field.

The authors in [8] introduce MediaPipe, a real-time framework that leverages convolutional networks for pose estimation on lightweight devices, such as smartphones and web platforms. MediaPipe's optimization for real-time applications makes it accessible for broader use.

The authors in [9] propose a hybrid framework combining convolutional networks with graphical models. By exploiting spatial relationships between body joints, the framework improves the accuracy of pose estimation significantly.

The authors in [10] present a unified system for lifting 3D pose estimation from a single image. Their method integrates 2D joint detection and 3D reconstruction in a single architecture, providing a comprehensive approach to pose estimation.

In the domain of real-time applications, the authors in [11] introduced MediaPipe, a multifaceted infrastructure designed for rapidity and efficacy. MediaPipe employs optimized convolutional networks to enable HPE on less capable devices, such as mobile and online platforms, hence expanding the practical application of advanced pose estimation methods.

The authors' invention of OpenPose in [12] signifies a notable progression in multi-person pose estimation. Utilizing Part Affinity Fields (PAFs) to delineate spatial associations among body parts, OpenPose facilitates real-time tracking of numerous persons in intricate settings, underscoring its effectiveness in dynamic contexts.

the comparison analysis in [13] evaluates the efficacy of various skeleton-based HPE libraries, such as OpenPose, PoseNet, MoveNet, and MediaPipe Pose. The paper demonstrates that MoveNet routinely surpasses other models in accuracy through extensive testing on picture and video datasets, especially in complicated movement scenarios, hence establishing important standards for future advancements in pose estimation technology.

The authors in [14] introduce Fast Pose Distillation (FPD), a strategy for training lightweight pose models with low computational cost, making it suitable for real-time applications.

Finally, the authors in [15] explore the use of end-to-end learning for HPE by combining a mixture of deformable parts with deep neural networks. The authors in [16] introduce Regional Multi-Person Pose Estimation (RMPE), a framework designed to handle bounding box inaccuracies in multi-person pose estimation.

2.2 Human Pose Estimation in Sports Activities

Given its promising features and capabilities, HPE is anticipated to be adopted extensively in sports, underscoring its crucial role in increasing athletic performance through accurate motion analysis and technique refinement. The authors of [2] emphasize the significance of HPE in sports, illustrating its ability to evaluate complex actions and enhance athlete performance. The promise of this technology transcends performance enhancement, encompassing vital applications in athlete training and injury prevention.

DOI: http://dx.doi.org/10.xxxxx/jistel.vXiX
e- ISSN xxxx-xxxx

The authors in [3] have utilized a 3D HPE framework for sporting activities, integrating temporal convolutions to capture the dynamic movements characteristic of sports, such as sprinting and jumping. Their scientific approach enhances understanding of athletic movements and increases the precision of real-time movement analysis in sports settings.

The study in [13] further investigates the capabilities of HPE by examining pose tracking algorithms tailored for sports. This study examines the differences in body posture and dynamic motions, improving the use of pose estimation technologies for thorough athletic performance evaluation and feedback.

The authors in [15] from the American Council of Exercise have delineated crucial principles for preserving proper posture throughout strength training movements, including squats, bicep curls, and deadlifts. These defined parameters are essential for evaluating movement precision and technique, acting as benchmarks that enhance training efficacy and mitigate injury risk.

The authors in [17] use RMPE to improve the accuracy of pose estimation in group settings, particularly applicable to sports activities involving multiple participants. By addressing bounding box inaccuracies and refining pose predictions, this framework is particularly useful for analyzing interactions and dynamic motions in team sports or group fitness sessions.

The study in [18] proposes detailed guidelines for evaluating movements in strength training, focusing on exercises such as squats, bicep curls, and deadlifts. These benchmarks serve as a foundation for both researchers and practitioners to assess the accuracy of techniques and to develop effective training programs that prioritize proper form.

The authors in [19] provide a comprehensive overview of deep learning applications in sports, categorizing methodologies into perception, understanding, and decision-making. This framework highlights the multifaceted role of deep learning in improving athletic performance by enabling better data-driven insights and tailored training regimens.

The authors in [20] explore pictorial structures for detecting and analyzing articulated poses in sports. Their robust framework is particularly useful for tracking body movements during dynamic activities, such as running or jumping, and provides a foundation for developing advanced monitoring systems for athletes in motion-intensive sports.

The authors in [21] detail posture and movement standards for strength training, emphasizing the importance of maintaining proper form to avoid injury. Their guidelines act as a critical reference for both individual athletes and trainers, ensuring that movements are performed efficiently and safely to achieve optimal outcomes.

The research in [22] presents the Exercise and Performance Learning Assistant System (EPLAS), developed on the Node.js framework to enhance HPE applications for yoga practices. This advanced system takes and assesses user poses in real-time, delivering prompt feedback that is crucial for enhancing yoga pose precision and overall performance.

Likewise, the authors in [23] advanced the EPLAS to assist users in home-based exercise or performance learning settings. This system employs Node.js to automatically assess and compare the alignment of body keypoints between a user and an instructor, specifically in the context of fitness exercises as referenced in source [22].

The findings in [24] expand the use of stance estimation to the identification of yoga poses, demonstrating the extensive applicability of HPE in sports and fitness activities. Finally, the study detailed in [25] utilizes these technical improvements in rehabilitation, where accurate pose estimate is essential for assessing and rectifying body movements to facilitate physical

DOI: http://dx.doi.org/10.xxxxx/jistel.vXiX
e- ISSN xxxx-xxxx

recovery. Human pose estimate serves as a transformative instrument in sports science, improving performance and safety across multiple athletic disciplines through its wide uses.

3. METHODS

In this study, primary data were collected to develop and evaluate a two-dimensional human pose estimation system using keypoint angle detection for basic strength exercises. The data consist of video recordings captured directly during strength exercises, namely squats, bicep curls, and deadlifts.

In this research, we collaborated with professional fitness trainers in performing the exercises. Video data were recorded using a standard webcam in an indoor controlled environment with adequate lighting to optimize image quality for pose estimation. The trainer performed 10 repetitions of each exercise. The recordings were made at 30 frames per second and a resolution of 640×480 pixels to balance data quality and computational requirements. No markers or special clothing were used to ensure that the system could operate in real-world conditions without specialized equipment.

After recording videos directly at the training site with a fitness expert, the ideal movement patterns were determined by analyzing the joint angles required for exercises such as squats, bicep curls, and deadlifts. Using the MediaPipe system, we calculated the angles involved in these movements (**Figure 1**). The recorded videos were evaluated to establish the optimal joint angles for each exercise, and with the guidance of the expert, thresholds were set for these movements. A margin of 20 degrees was determined to allow the system to accommodate slight variations in user movements while still recognizing the correct motion patterns. These thresholds were applied to quantify the number of correct repetitions during exercise sessions, enabling users to identify any mistakes in their movements.

3.1. Hardware Specification

Two devices were utilized in this study: a webcam and a personal computer (PC) functioning as the server. The webcam was connected to the PC, which accessed a custom-built website capable of capturing user posture and movements in real-time. The website also allowed for the recording of user movements, enabling the review of session results after training. The real-time camera transmitted data to the server, where the video was processed to extract the X, Y, and Z coordinates of 33 key reference points. These coordinates are then used to calculate eight key joint angles, specifically the joints of the elbows, shoulders, hips, and knees. **Tables 1** and 2 provide the specifications the PC server and camera utilized in this study.

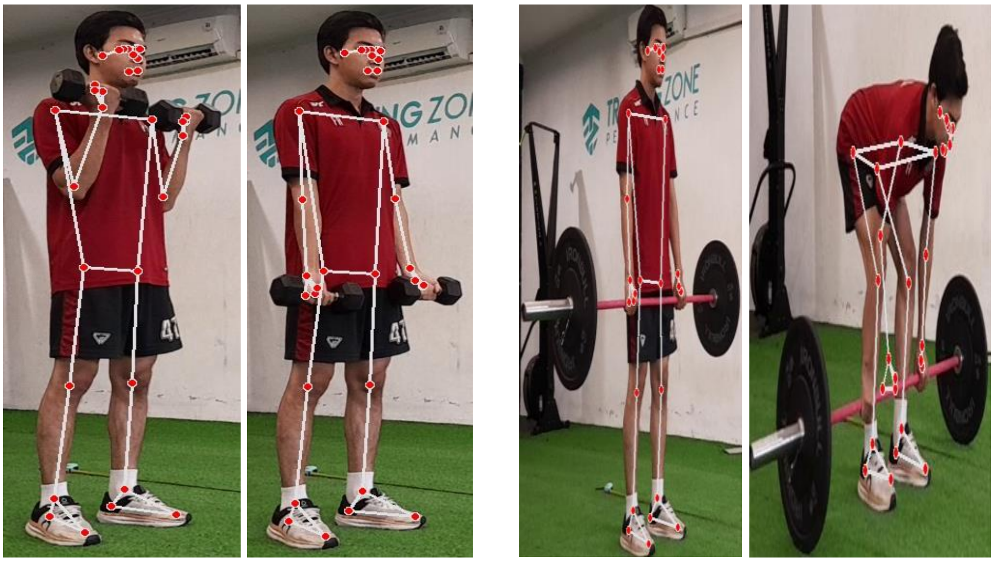
Table 1. Computer Specification

Atribute	Specification
Manufacturer	ASUS
Model	A442U
Camera Internal	VGA Web Camera
CPU	Intel® Core™ i5 8250U Processor (6M Cache, up to 3.40 GHz)
System Operation	Windows 10 Home
RAM	12GB DDR4 2133MHz SDRAM
SSD	Samsung SATA III 870 500GB

DOI: http://dx.doi.org/10.xxxxx/jistel.vXiX
e- ISSN xxxx-xxxx

Table 2. Camera External Specification

Atribute	Specification
Manufactirer	Micopack Webcam
Camera Physical Pixel	1 MP, 720p HD
Camera Angle	60 degrees
Interface type	USB2.0 + 3.5mm jack for MIC, 145CM



(a) (b)

Figure 1. Bicep Curl (a) and Deadlift (b) images processed using Mediapipe

3.2 System Design

The system's design used in this study are illustrated in **Figure 2**, detailing the testing process of a two-dimensional pose estimation system using angle detection for basic strength training. The steps of the research methodology are as follows:

1. Equipment and Experimental Setup
- Hardware:
  a) A personal computer (PC) was used as the server for data processing.
  b) A webcam connected to the PC was used to record user movements.
  c) The camera was positioned before the user at a height of 46cm from the floor to ensure the entire body was visible. The camera placement is illustrated in **Figure 3**.
- Software:
  a) A custom-built website was designed to capture, and process pose data in real-time.
  b) A framework based on MediaPipe was utilized for extracting the body reference points (landmarks).

2. Experimental Procedure
   a) The camera was fixed in front of the user, ensuring the field of view included the user's entire body.
   b) The PC was configured to run the server that handled pose data processing in real-time.

DOI: http://dx.doi.org/10.xxxxx/jistel.vXiX
e- ISSN xxxx-xxxx

c) The user accessed the website via a browser on the PC and was instructed to stand in front of the camera.
d) The program provided direct feedback to the user, guiding them to adjust their orientation and distance until the entire body was detected as in **Figure 3**.
e) Once the correct distance and orientation were achieved, the user performed a series of strength exercises, such as bicep curls, squats, and deadlifts, according to their capability.
f) The camera recorded each movement, and the server processed the video to extract the X, Y, and Z coordinates of 33 key body reference points.
g) The server analyzed the number of correct movements performed during the exercises based on the angle data generated.
h) The analysis results were displayed on the website interface in real-time as graphical and numerical visualizations.

3. System's Process Diagram

Figure 2 illustrates the process flow diagram of the system, which includes:

a) Input Data: The camera captures the user's video.
b) Data Processing: The video is transmitted to the server for coordinate extraction and angle calculation.
c) Analysis: The system evaluates movement accuracy and provides feedback.
d) Output: Correct movements are displayed to the user via the website interface.

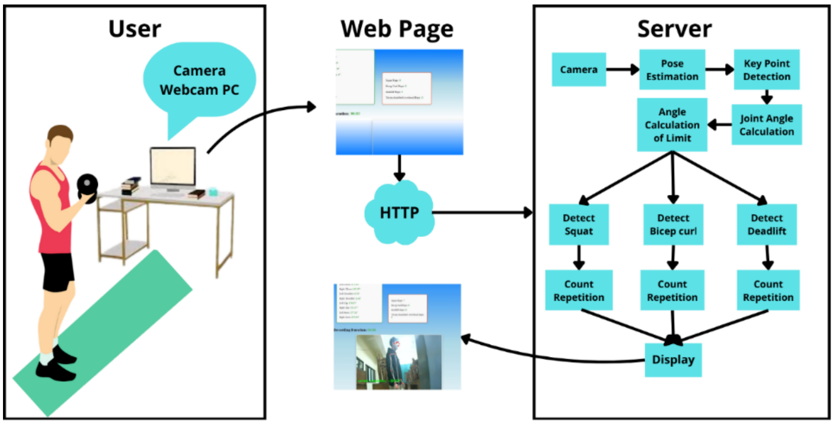


Figure 2. An overview of the method used in this experiment.

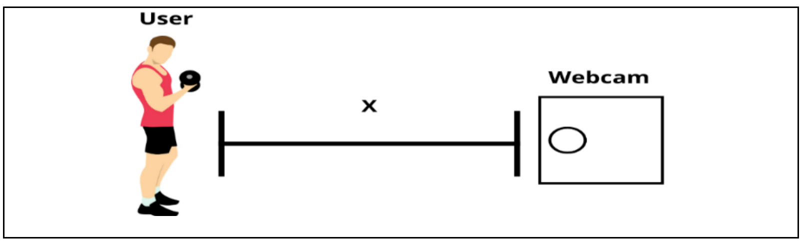


Figure 3. Distance Setting on Camera.

3.3 Repetition Counter Feature

MediaPipe enables the extraction of 33 landmark coordinates from photos/videos containing the human body. **Figure 4**. illustrates the 33 landmarks extracted by mediapipe. In this study, the coordinates of the landmarks were taken from an instructor video and a real-time user video. The extraction of the instructor's 33 landmarks coordinates was taken beforehand using a

DOI: http://dx.doi.org/10.xxxxx/jistel.vXiX
e- ISSN xxxx-xxxx

camera with specifications in **Table 2**. The coordinates will then be used to calculate 8 main joint angles from the user's body. The user's 8 main joint angles will be compared with the instructor's 8 main joint angles of the same strength training action stage to evaluate the accuracy of the user's reaction.

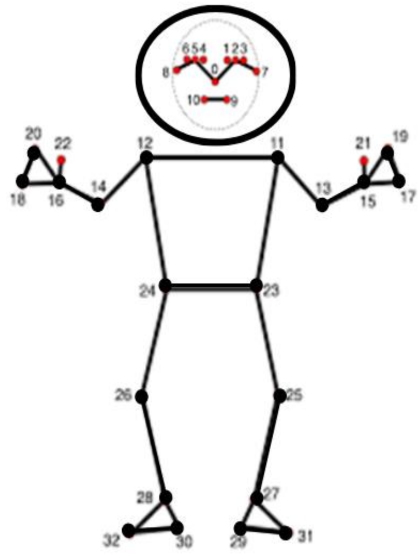


Figure 4. Keypoints extracted from Mediapipe

Calculation of joint angles from extracted landmarks was done using the following formula:

$$\theta = cos^{-1}\left(\frac{ba * bc}{\|ba\| * \ \|bc\|}\right) \tag{3}$$

a, b and *c* refer to 3 landmarks consist of x and y coordinate, *ba*bc* is the product of vector *ba* and *bc,* while $\|ba\|$ * $\|bc\|$ refers to the norms of vector *ba* and *bc* respectively.

3.3.1 Adjust System Coordinates on Camera

Comparing the extracted 8 main joint angles can be difficult as the coordinate systems of the two photos/videos may differ due to differences in camera architecture or distance between the subject and camera. Therefore, the user's coordinate system has to be adjusted to be coincident with the instructor's coordinate system.

This study took the approach of adjusting 8 main joint angles through a shoulder-length comparison ratio,

$$\theta' = \theta\left(\frac{\|\mathrm{RP1} - \mathrm{RP2}\|}{\|\mathrm{UP1} - \mathrm{RP2}\|}\right) \tag{4}$$

to minimize the difference between the instructor's and user's strength training action stage. θ' refers to the normalized joint angle obtained from the dot product of the user's joint angle and the shoulder length comparison ratio. $\|RP1 - RP2\|$ And $\|UP1 - UP2\|$ refer to the instructor's and user's shoulder length; these are obtained by applying Euclidean Distance to their landmark respectively.

The shoulder proportion value is used in the "Proportion Comparison" attribute and obtained through Percentage Formula as:

$$PR = \left(\frac{\|\mathrm{RP1} - \mathrm{RP2}\| - \ \|\mathrm{UP1} - \mathrm{UP2}\|}{\|\mathrm{RP1} - \mathrm{RP2}\|}\right)100 \tag{5}$$

DOI: http://dx.doi.org/10.xxxxx/jistel.vXiX
e- ISSN xxxx-xxxx

The shoulder proportion threshold is set at 10 percent. A percentage value above 10 percent indicates that the user's shoulder width is shorter than the instructor's shoulder width, which occurs because the user is positioned too far from the camera. If the user's shoulder proportion falls within the 10 percent range, deemed as the "Correct Orientation," the system will display the "Angle List" attribute containing the eight key joint angles adjusted for analysis.

Each movement, such as squats, bicep curls, and deadlifts, is assigned a maximum allowable angle deviation threshold of 20 degrees for the eight primary joint angles. This threshold is implemented to reduce the risk of injury during independent training and allows users to evaluate errors immediately.

Every strength training exercise is divided into two stages: the "Down" phase (initial stage) and the "Up" phase (final stage). Movements in these phases are considered correct when the eight key joint angles adjusted by the user closely match or are identical to those of the instructor. For example, in the "Down" phase of a squat, the eight threshold angle values include a knee joint angle deviation of less than 20 degrees and a range between 150 and 170 degrees for knee, hip, and elbow joint angles. A repetition is recorded when both phases are successfully completed by the user, indicated by an increment in the "Counter" attribute.

4. RESULTS AND DISCUSSION

4.1. Results

Tables 3, 4, and **5** present the results of pose estimation, the counter feature, and the repetition evaluation conducted by the instructor for each user. Users performed movements randomly in front of the camera at the same distance in meters. These tables provide data on the age, gender, height (H), weight (W), distance in meters, the number of repetitions performed by the user (Counted Repetition), and the number of repetitions validated by the server (Validated Repetition).

From the various movements performed by users, which were conducted randomly or out of sequence during basic strength exercises such as bicep curls, squats, and deadlifts, the results are shown in **Table 3**, **Table 4**, and **Table 5**. These tables display the number of movements successfully performed by users.

- In **Table 3**, the average score is 5 points.
- In **Table 4**, the average score is 8.5 points.
- In **Table 5**, the average score is 10 points.

When compared with the angles recorded by the instructor, the system validates the number of correct movements as follows:

- **Table 3** shows a comparison score of 1 point.
- **Table 4** shows a comparison score of 1.5 points.
- **Table 5** shows a comparison score of 2 points.

Table 3. Deadlift Results

ID	Gender	Age	H(m)	W(Kg)	Distance (m)	Counted Repetition	Validated Repetition
1	Male	12	1.46	48	1.8	5	3
2	Male	22	1.7	59	2.2	5	4
3	Male	27	1.75	61	2.4	5	4
4	Male	56	1.76	68	2.3	5	5
Mean						5	4

DOI: http://dx.doi.org/10.xxxxx/jistel.vXiX
e- ISSN xxxx-xxxx

Table 4. Bicep Curl Results

ID	Gender	Age	H(m)	W(Kg)	Distance (m)	Counted Repetition	Validated Repetition
1	Male	12	1.46	48	1.8	6	4
2	Male	22	1.7	59	2.2	10	8
3	Male	27	1.75	61	2.4	10	9
4	Male	56	1.76	68	2.3	10	9
Mean						8.5	7

Table 5. Squat Results

ID	Gender	Age	H(m)	W(Kg)	Distance (m)	Counted Repetition	Validated Repetition
1	Male	12	1.46	48	1.8	10	7
2	Male	22	1.7	59	2.2	10	8
3	Male	27	1.75	61	2.4	10	9
4	Male	56	1.76	68	2.3	10	8
Mean						10	8

These results may be influenced by several factors. First, the discrepancy in shoulder proportions between the users and the instructor may prevent the system from accurately reading the movements in a short time, particularly at the start or completion of the motion. Second, differences in the angles produced by users compared to the instructor may arise due to the fact that the users tested have limited experience in performing strength exercises, such as bicep curls, squats, or deadlifts. This lack of familiarity with strength training techniques may lead to improper form, resulting in reduced validation scores by the system.

4.1. Evaluation Discussion

To evaluate performance, matrix calculations are required, as shown in Equation (6) to determine accuracy and Equation (7) to determine the system's sensitivity. These calculations utilize the results from CR (Count Repetition) and VR (Validated Repetition) to derive the values for True Positive (TP), False Positive (FP), and False Negative (FN). The formulas are as follows:

$$\text{Accuracy} = \frac{TP}{TP + FP + FN} \tag{6}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{7}$$

The substitution results presented in **Table 6** for the Deadlift movement show an accuracy of 80% with 4 false positives. In **Table 7**, for the Bicep Curl movement, the accuracy achieved is 82.8% with 6 false positives. Meanwhile, in **Table 8**, for the Squat movement, the system demonstrates an accuracy of 80% with 8 false positives.

Table 6. Deadlift Performance Matric

No.	CR	VR	TP	FP	FN
1	5	3	3	2	0
2	5	4	4	1	0
3	5	4	4	1	0
4	5	5	5	0	0

DOI: http://dx.doi.org/10.xxxxx/jistel.vXiX
e- ISSN xxxx-xxxx

Each of these tables uses a sensitivity of 100%, indicating that the system based on MediaPipe can be effectively utilized for independent exercise activities, providing reliable movement validation and performance feedback.

Table 7. Bicep Curl Performance Matric

No.	CR	VR	TP	FP	FN
1	6	4	4	2	0
2	10	8	8	2	0
3	10	9	9	1	0
4	10	9	9	1	0

Table 8. Squat Performance Matric

No.	CR	VR	TP	FP	FN
1	10	7	7	3	0
2	10	8	8	2	0
3	10	9	9	1	0
4	10	8	8	2	0

To validate the proposed Counted Repetition feature, this study adopts the approach of applying the Pearson Correlation Coefficient to CR and VR among 4 users.

$$\mathrm{r} = \frac{n(xy) - (x)(y)}{\sqrt{(n\sum(x^2) - \sum(x)^2)\,(n\sum(y^2) - \sum(y)^2)}} \tag{6}$$

is used to find the correlation between CR and VR. **Table 6, 7,** and **8** shows the value and sum of x, y, xy, x^2, and y^2 for **table 3, 4,** and **5**; the value of x and y refers to CR and VR.

Table 9. Deadlift Correlation Calculation

ID	x	y	xy	x^2	y^2
1	5	3	15	25	9
2	5	4	20	25	16
3	5	4	20	25	16
4	5	5	25	25	25
Sum	25	16	80	100	66

Table 10. Bicep Curl Correlation Calculation

ID	x	y	xy	x^2	y^2
1	6	4	24	36	16
2	10	8	80	100	64
3	10	9	90	100	81
4	10	9	90	100	81
Sum	36	30	284	336	242

Table 11. Squat Correlation Calculation

ID	x	y	xy	x^2	y^2
1	10	7	70	100	49
2	10	8	80	100	64
3	10	9	90	100	81
4	10	8	80	100	64

DOI: http://dx.doi.org/10.xxxxx/jistel.vXiX
e- ISSN xxxx-xxxx

Sum	40	32	320	400	258

Due to the lack of variation in x values in **Tables 9** and **11**, the Pearson correlation coefficient cannot be determined. In contrast, **Table 10** demonstrates a highly accurate correlation coefficient of 0.98, as only one x value differs, contributing to the measurable linear relationship.

5. CONCLUSION

This paper proposes a two-dimensional human pose estimation system for basic strength training exercises to assist individuals in performing strength training independently and provide a solution to prevent users from sustaining injuries during solo workouts. The results obtained from the testing phase were evaluated by an instructor and then compared.

Based on the system's performance analysis, it can be concluded that the system is reasonably accurate, achieving an average accuracy of 80%. Additionally, a strong correlation factor of 0.98 confirms the effectiveness of the proposed system. It is necessary to add more types of strength exercises, as well as a feature for automatically detecting and selecting exercise types based on the user's current pose. Furthermore, implementation on mobile devices is required and will be included in future work.

6. ACKNOWLEDGMENT

This research is sponsored by the *Institute for Research and Community Service (LPPM) of Universitas Negeri Surabaya*. This research is also in partnered *with Okayama University, Japan,* under the Memorandum of Agreement with Faculty of Engineering, UNESA. The data acquisition was collected under the partnership agreement with *Training Zone ID.*

7. AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. The authors confirmed that the paper was free of plagiarism.

8. AUTHORS' CONTRIBUTION

Achmad Ivan Taruna Jaya: conceptualization, data collection, formal analysis, investigation, methodology, visualization, writing-original draft. Pradini Puspitaningayu: conceptualization, supervision, validation, writing-review and editing, funding acquisition. Athaya Pradipa Adhiwangsa: software, project administration, resources, data curation, investigation. Nobuo Funabiki: supervision, validation.

9. REFERENCES

[1] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Comput. Vis. Image Underst.*, vol. 104, no. 2-3 SPEC. ISS., pp. 90–126, 2006, doi: 10.1016/j.cviu.2006.08.002.

[2] Y. Chen, Y. Tian, and M. He, "Monocular human pose estimation: A survey of deep learning-based methods," *Comput. Vis. Image Underst.*, vol. 192, no. December 2019, p. 102897, 2020, doi: 10.1016/j.cviu.2019.102897.

[3] D. Pavllo, C. Feichtenhofer, D. Grangier, and M. Auli, "VideoPose3D," *Proc. IEEE*

Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2019-June, pp. 7745–7754, 2019.

[4] V. Belagiannis and A. Zisserman, "Recurrent Human Pose Estimation," *Proc. - 12th IEEE Int. Conf. Autom. Face Gesture Recognition, FG 2017 - 1st Int. Work. Adapt. Shot Learn. Gesture Underst. Prod. ASL4GUP 2017, Biometrics Wild, Bwild 2017, Heterog. Face Recognition, HFR 2017, Jt. Chall. Domin. Complement. Emot. Recognit. Using Micro Emot. Featur. Head-Pose Estim. DCER HPE 2017 3rd Facial Expr. Recognit. Anal. Challenge, FERA 2017*, pp. 468–475, 2017, doi: 10.1109/FG.2017.64.

[5] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9912 LNCS, pp. 483–499, 2016, doi: 10.1007/978-3-319-46484-8_29.

[6] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2D human pose estimation: New benchmark and state of the art analysis," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 3686–3693, 2014, doi: 10.1109/CVPR.2014.471.

[7] S. Kulkarni, S. Deshmukh, F. Fernandes, A. Patil, and V. Jabade, "PoseAnalyser: A Survey on Human Pose Estimation," *SN Comput. Sci.*, vol. 4, no. 2, pp. 1–11, 2023, doi: 10.1007/s42979-022-01567-2.

[8] A. Toshev and C. Szegedy, "DeepPose_Human_Pose_2014_CVPR_paper.pdf," *2014 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1653–1660, 2014, [Online]. Available: https://www.cv-foundation.org/openaccess/content_cvpr_2014/html/Toshev_DeepPose_Human_Pose_2014_CVPR_paper.html

[9] J. Tompson, A. Jain, Y. LeCun, and C. Bregler, "Joint training of a convolutional network and a graphical model for human pose estimation," *Adv. Neural Inf. Process. Syst.*, vol. 2, no. January, pp. 1799–1807, 2014.

[10] D. Tome, C. Russell, and L. Agapito, "Lifting from the deep: Convolutional 3D pose estimation from a single image," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 5689–5698, 2017, doi: 10.1109/CVPR.2017.603.

[11] S.-E. Wei, V. Ramakrishna, T. Kanada, and Y. Sheikh, "Pose Machines :," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016.

[12] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 5686–5696, 2019, doi: 10.1109/CVPR.2019.00584.

[13] J. L. Chung, L. Y. Ong, and M. C. Leow, "Comparative Analysis of Skeleton-Based Human Pose Estimation," *Futur. Internet*, vol. 14, no. 12, 2022, doi: 10.3390/fi14120380.

[14] Z. Feng, X. Zhu, and M. Ye, "Fast Pose Estimation," *Comput. Vis. Pattern Recognit.*, pp. 3517–3526, 2019.

[15] C. Lugaresi *et al.*, "MediaPipe: A Framework for Building Perception Pipelines," 2019, [Online]. Available: http://arxiv.org/abs/1906.08172

[16] W. R. Czart and S. Robaszkiewicz, "Superconducting properties of the η-pairing state in the Penson - Kolb - Hubbard model," *Acta Phys. Pol. A*, vol. 106, no. 5, pp. 709–713, 2004, doi: 10.12693/APhysPolA.106.709.

[17] H. Fang, S. Xie, Y. Tai, and C. Lu, "RMPE : Regional Multi-Person Pose Estimation".

[18] W. Yang, W. Ouyang, H. Li, and X. Wang, "End-to-End Learning of Deformable Mixture of Parts and Deep Convolutional Neural Networks for Human Pose Estimation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-December, pp. 3073–3082, 2016, doi: 10.1109/CVPR.2016.335.

[19] Z. Zhao *et al.*, "A Survey of Deep Learning in Sports Applications: Perception,

Comprehension, and Decision," pp. 1–20, 2023, [Online]. Available: http://arxiv.org/abs/2307.03353

[20] M. Andriluka, S. Roth, and B. Schiele, "Pictorial structures revisited: People detection and articulated pose estimation," *2009 IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2009*, pp. 1014–1021, 2009, doi: 10.1109/CVPRW.2009.5206754.

[21] L. C. Dalleck, A. M. Dalleck, and B. R. Byrd, "Effectiveness of the ACE Mover Method to Elicit Positive Healthy Behavior and Lifestyle Changes," vol. 16, pp. 19–25, 2020.

[22] T. Anggraini, P. Puspitaningayu, N. Funabiki, S. -W. Shen, W. -C. Huang and C. -P. Fan, "An Implementation of Exercise and Performance Learning Assistant System Platform Using Node.js," 2022 IEEE International Conference on Consumer Electronics - Taiwan, Taipei, Taiwan, 2022, pp. 193-194, doi: 10.1109/ICCE-Taiwan55306.2022.9869003.

[23] Tri Anggraini, N. Funabiki, P. Puspitaningayu, S. -W. Shen, W. -C. Huang and C. -P. Fan, "Implementation and Evaluation of Exercise and Performance Learning Assistant System Platform for Yoga Pose Practices Using Node.js," 2022 Fifth International Conference on Vocational Education and Electrical Engineering (ICVEE), Surabaya, Indonesia, 2022, pp. 100-105, doi: 10.1109/ICVEE57061.2022.9930470.

[24] A. Wang, G. Chen, C. Shang, M. Zhang, and L. Liu, "Web-Age Information Management," vol. 7923, pp. 29–40, 2013, doi: 10.1007/978-3-642-38562-9.

[25] A. Raza, A. M. Qadri, I. Akhtar, N. A. Samee, and M. Alabdulhafith, "LogRF: An Approach to Human Pose Estimation Using Skeleton Landmarks for Physiotherapy Fitness Exercise Correction," *IEEE Access*, vol. 11, no. October, pp. 107930–107939, 2023, doi: 10.1109/ACCESS.2023.3320144.

DOI: http://dx.doi.org/10.xxxxx/jistel.vXiX
e- ISSN xxxx-xxxx