Vision-Based Posture Classification of the Standard Squat Exercise Using MediaPipe Pose

Renzo John L. Virtucio
and Prospero C. Naval, Jr.
Computer Vision & Machine Intelligence Group
Department of Computer Science
College of Engineering
University of the Philippines-Diliman

ABSTRACT

Maintaining an active lifestyle plays a huge role in caring for one's overall health and well-being. When engaging in physical exercise, it is highly important to ensure proper execution of the exercise for safety and the maximization of its benefits, and this is usually attainable through the supervision of those with experience and knowledge in the field of physical fitness such as professional trainers. However, the access to professional help has grown scarce during the pandemic, thus a cause for concern is the lack of education and supervision for novices in physical exercise and those who can only exercise at home. To promote, proper, safe, and effective exercise, the researcher developed a vision-based solution to classifying postures and providing feedback during the performance of resistance training exercise. Due to time constraints and the lack of a comprehensive dataset which spans various resistance training exercises, the researcher had to limit the scope of the resistance training exercises to only the standard squat. The researcher gathered videos of squat sets from 10 conveniently sampled individuals, resulting in a total of 60 raw videos across the 6 identified standard squat posture classes. The researcher developed a Support Vector Machine classifier which obtained an accuracy score of 92.92%, precision scores of no less than 81%, recall scores of no less than 83%, and f1-scores of no less than 86% across all 6 squat posture classes plus the newly introduced standing posture class.

1. INTRODUCTION

1.1 Statement of the Problem

The importance of caring for their health and well-being, as well as the various ways to achieve a lifestyle which promotes these, is predominantly taught to individuals beginning at a very young age. Alongside having a healthy diet, engaging in physical activity or exercise is fundamental in maintaining or improving one's health condition. Although aerobic exercise, or more commonly known as "cardio", may still be the more advocated and more prescribed mode of exercise, resistance training has been gaining popularity as more have been learned about the advantages and health benefits it provides.

Resistance training is defined to be a category of physical activity that engages the muscles to work against an additional force or weight (also referred to as resistance) [1]. Resistance training exercises include lifting weights such as dumbbells and barbells, and even body weight exercises such as sit-ups and push-ups. The main advantage of resistance training over aerobic exercise is its high effectiveness in the development of muscular strength and muscular mass albeit this is not its only benefit. Regular resistance training has also been observed to increase bone strength and bone density which helps in the prevention of fall fractures [1][2]. Resistance training may also reduce the risk for heart disease as it is known to lower body fat, decrease blood pressure, and improve cholesterol level [1]. Furthermore, resistance training not only affects physical health for research evidences show that it can lessen anxiety symptoms, improve cognition among older adults and reduce symptoms of depression among diagnosed patients [3]. Resistance training also promotes independence and improves self-esteem among individuals [1][2][3].

Along with its health benefits, resistance training is also accompanied by risk of injury. A 20-year survey on weight training injury trends reveal that an estimate of about 980,173 weight training-related injuries occurred in the United States from 1978 through 1998 [4]. Another study conducted among recreational weightlifters in Nnewi, Nigeria show a prevalence rate of 47.3% [5]. The types of reported injury vary from soft-tissue injury, fracture or dislocation, laceration, and general inflammation and pain [4][5]. To promote safety while engaging in resistance training and reduce the occurrence of injuries, medical institutions and researchers alike have established guidelines and recommendations. Although the guidelines and recommendations may vary from one author to another, they all emphasize the same aspects of safe and effective training, namely education on proper techniques, qualified supervision, and an appropriate training program [4][6][7].

1.2 Significance of this Study

The COVID-19 pandemic has drastically affected physical activity as for many individuals, they are unable to continue their form of exercise due to the closure of exercise facilities such as gyms, pools, and stadiums. A recent study on physical activity at home during the pandemic in Saudi Arabia shows a significant decrease (57.1%) between the time spent participating in physical activity before and during the COVID-19 lockdown [8]. Among those who used to engage in regular exercise before the pandemic, while there are some who were able to find alternative forms of exercise, the others were not able to adjust accordingly due to loss of motivation which may be attributed to a distorted social context [8][9][10]. Furthermore, since the access to professional help has grown scarce, another cause for concern is the lack of education and supervision for those who are forced to exercise at home which may contribute to the already high number of resistance training injuries that occurred at home [4]. Therefore, the researcher aimed to develop a tool that promotes proper, safe, and effective exercise by being able to classify and provide feedback on the posture during the performance of a resistance training exercise. This tool may be proven to be useful in educating novices in resistance training and in motivating individuals who want to get back into an active lifestyle.

1.3 Scope of this Study

Multiple studies have shown the effective application of artificial intelligence in the field of physical fitness, and past researchers attempted to address problems within this field through various approaches and techniques. A clear factor which divides the existing literature is the method or technology used to quantify their data, and the researcher identified two prevalent technologies: wearable sensors and computer vision. In consideration of available resources, the researcher decided to employ a vision-based solution to classifying postures during the performance of resistance training exercise. Furthermore, due to the lack of a comprehensive dataset which spans various resistance training exercises, the researcher decided to build their own dataset for this project. To be able to have time for both gathering data and developing the classifying model, the researcher decided to focus on only one form of resistance training exercise, which is the standard squat. A standard squat is performed with the following steps [11]:

1. Stand tall with your feet hip distance apart. Your hips, knees, and toes should all be facing forward.

2. Bend your knees and extend your buttocks backward as if you are going to sit back into a chair. Make sure that you keep your knees behind your toes and your weight in your heels.

3. Rise back up and repeat.

2. RELATED WORK

2.1 Lightweight Machine Learning-Based Approach for Supervision of Fitness Workout

Depari *et al.* [12] developed a lightweight system for the supervision of fitness workout with the use of Inertial Measurement Units (IMUs) embedded with an accelerometer and a gyroscope. Their system used a supervised Latent Dirichlet Allocation (LDA), claimed to be more efficient with computational resources than Support Vector Machines, for exercise classification, and they also included a repetition counter by detecting peaks of the velocity signal. The exercise classifier ensures an accuracy of over 93% although there was great variability in terms of its recall and precision, while their repetition counter shows a maximum average error of 6%.

2.2 Automatic Classification of Squat Posture Using Inertial Sensors: Deep Learning Approach

A study which addressed a more specific problem, one that the researcher aims to tackle as well, was conducted by Lee *et al.* [13] as they focused on the classification of squat posture. They also utilized IMUs, 5 for each participant, to quantify the squats. The study focused on 6 classifications of squat posture with 5 being incorrect forms and 1 being the correct form as seen in Figure 1. They developed 2 classifiers, a Random Forest and a Convolutional Neural Network-Long Short Term Memory (CNN-LSTM), and compared their accuracies. They obtained an accuracy of 91.7% for the CNN-LSTM, and an accuracy of 75.4% for the Random Forest.

Squat	Description	Figure	Squat	Description	Figure
Acceptable (ACC)	Normal squat		Knee varus (KVR)	Both knees pointing outside during exercise	
Anterior knee (AK)	Knees ahead of toes during exercise		Half squat (HS)	Insufficient squatting depth	
Knee valgus (KVG)	Both knees pointing inside during exercise		Bent over (BO)	Excessive flexing of hip and torso	

Figure 1: Squat Posture Classes from Lee *et al.*

2.3 Temporal Distance Matrices for Squat Classification

Although the use of wearable sensors shows great potential in quantifying exercises, the researcher recognizes computer vision to be another effective solution for this task. Instead of using accelerometers and gyroscopes to measure movement, computer vision uses images and videos, which is arguably more accessible since many, if not the majority, own a camera in one form or another. Multiple studies show that a squat classification system similar to that of Lee *et al.* [13] can be developed using the technology of computer vision instead. Ogata *et al.* [14] identified 7 squat classes and was able to create a squat classifier using a Convolutional Neural Network (CNN) with 1 convolution on temporal distance matrices derived from the output of a 3D pose detector. Their classifier obtained an accuracy of 81.05% on their own dataset which features only 1 individual, 88.93% on their own dataset which features multiple individuals, and 78.26% on their dataset built from YouTube videos.

2.4 Design of Mobile Personal Workout Assistant Using Deep Learning

Park [15] also developed a squat classifier of their own as a component of their mobile personal workout assistant. Their classifier was a model with both temporal convolution layers and LSTM blocks which can only classify a squat as either correct or incorrect on 2D data extracted using OpenPose. Park obtained a test accuracy of 84.98%.

3. METHODOLOGY

The developed vision-based solution, depicted in 2 for classifying standard squat posture involves selecting frames from videos of individuals performing the standard squat exercise for each posture class, extracting coordinates of body landmarks from the selected frames, deriving new features from the body landmark coordinates, and training a learning model. To have a benchmark for the vision-based solution, the researcher decided to adopt the 6 squat posture classes that Lee *et al.* identified in their study as seen in Figure 1.

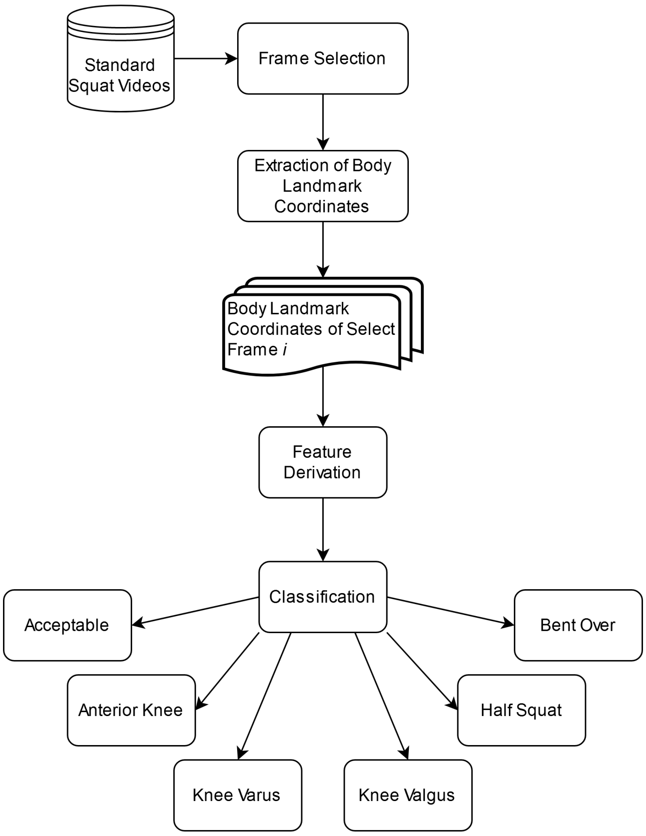

Figure 2: Developed Vision-Based Solution

3.1 Data Collection

The researcher gathered data from 10 individuals who were conveniently sampled. Each participant was asked to provide 6 video recordings of themselves performing 5 repetitions for each of the 6 squat posture classes. The researcher gathered a total of 60 raw videos across the 6 standard squat posture classes, which is equivalent to 50 repetitions per squat posture class. A new class, encoded as 'others', was introduced to encompass the subject's standing posture before and after performing the standard squat, and prevent the classifying model to make predictions when the subject is just in a standing position.

3.2 Data Preprocessing

Each raw video of 5 repetitions was split into five (5) individual videos, one (1) for each repetition. Frames for each repetition video were extracted and filtered such that only the frames which visibly demonstrate the squat posture classes remained. MediaPipe Pose [16] was then used to extract data from the selected frames. Data extracted were the x, y, z coordinates and the visibility of each of thirty-three (33) body landmarks.

3.3 Deriving Features

The extracted body landmark data was loaded into individual dataframes, then concatenated to form a main dataframe. New columns, which describe the angles formed by certain joints, and the distances between joints, were computed and derived from the x, y, z coordinates columns then added to the dataframe. The original columns, i.e., the landmark coordinates and visibility, were dropped. After preprocessing, the main dataframe consisted of 11,022 rows (or datapoints) and 22 columns (inclusive of the features and the target label). Its frequency distribution can be seen in Table 1.

	Count
Acceptable	1388
Anterior Knee	1430
Bent Over	1402
Half Squat	1726
Knee Valgus	1348
Knee Varus	1328
Standing	2400

Table 1: Frequency Distribution

3.4 Classification Model

The data was split 75/25 for training and testing respectively. The researcher ensured that the participants and their videos assigned to training are mutually exclusive to those assigned to testing so as to contribute to the robustness of the learning model. The researcher opted for a Support Vector Machine classifier due to the general low usage of computational resources by traditional machine learning models compared to deep learning models. The Support Vector Machine classifier was trained with the following major parameters:

- Kernel: Radial Basis Function (RBF)
- C: 1
- Gamma: 0.001
- Decision Function: One-vs-rest

4. EXPERIMENTS

The Support Vector Machine classifier obtained an accuracy score of 0.9292101341281669 or 92.92%. Its performance for each of the squat posture classes is described by its confusion matrix shown in Figure 3 and its classification report shown in Table 2. Take note of the following codes used to denote the squat posture classes:

- acc (Acceptable)

- ak (Anterior Knees)
- bo (Bent Over)
- hs (Half Squat)
- kvg (Knee Valgus)
- kvr (Knee Varus)
- others (Standing)

First and foremost, it can be observed that the model obtained a perfect precision, recall and f1-score for the Standing posture class, which was newly introduced for this project. This indicates that for the test set, the model was able to identify without fail when the subject was in a standing position, and this may be attributed to the fact that the standing posture could be easily distinguished from the rest of the posture classes.

For the original 6 posture classes, it can be observed that the model obtained the highest precision score for the Bent Over and Knee Valgus posture classes at 100%, while its precision score is the lowest for the Acceptable posture class at 82%. As for its recall score, the highest was obtained for the Acceptable posture classes at 100%, and the lowest for the Half Squat posture class at 84%. Finally, the model obtained the highest F1-score for the Bent Over posture class at 99% while its F1-score is the lowest for the Half Squat at 87%.

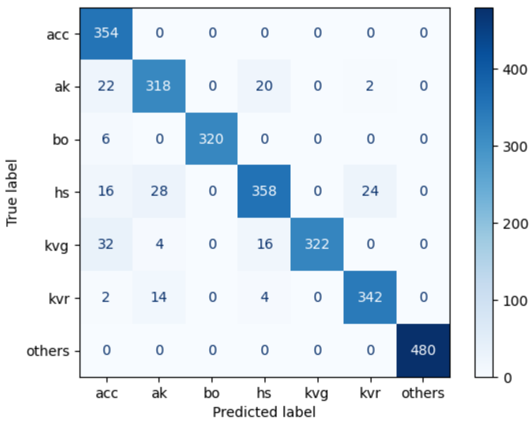

Figure 3: Confusion Matrix

	Precision	Recall	F1-score	Support
Acceptable	0.82	1.00	0.90	354
Anterior Knee	0.87	0.88	0.88	362
Bent Over	1.00	0.98	0.99	326
Half Squat	0.90	0.84	0.87	426
Knee Valgus	1.00	0.86	0.93	374
Knee Varus	0.93	0.94	0.94	362
Standing	1.00	1.00	1.00	480

Table 2: Classification Report

Compared to the classifier of Ogata *et al.* [14] which also used human pose estimation to quantify the squat exercise, the Support Vector Machine classifier developed in this study has a higher accuracy: 92.92% vs. 81.05%. It is important to consider, however, that Ogata *et al.* used different squat posture classes, and used a Convolutional Neural Network (CNN) as its classifier.

On the other hand, compared to the classifier of Lee *et al.* [13] which used inertial measurement units (IMUs) instead to quantify the squat exercise, the classifier developed in this study has a slightly higher accuracy in classifying across the same 6 squat postures: 92.92% vs. 91.7%. It is important to note as well that Lee *et al.* used a Convolutional Neural Network-Long Short Term Memory (CNN-LSTM) system as their classifier.

In spite of the classifier's performance, the researcher identified the following limitations:

1. The classifier is trained to classify in a fixed context. The following are assumed:
   - There is only 1 subject in the frame.
   - The subject is performing a standard squat.
   - The subject is either facing 45 degrees to the left or to the right with respect to the camera.
   - The camera is at the subject's torso level and captures the subject's entire body (from head to toe).
2. The following system PC specifications are recommended to ensure the intended performance of MediaPipe Pose and the developed classifier:
   - OS: 64-bit Windows 7, Windows 8.1, Windows 10.
   - Memory: 8 GB RAM
   - CPU: Intel i3-4150
   - Webcam Resolution: 640x480
3. The classifier's accuracy score, precision scores, recall scores and f1-scores could not be used as definitive indication of its robustness and overall performance in a real-world application especially because of the small sample from which the data used for this project was gathered.

The researcher developed a simple application prototype which utilizes the developed Support Vector Machine classifier to provide real-time feedback of the user's squat posture as they perform the exercise as shown in Figure 4.

5. CONCLUSION

To contribute to the promotion of proper, safe and effective exercise, the researcher developed a Support Vector Machine model that is able to classify squat postures based on the coordinates of body landmark data extracted through MediaPipe Pose, an established machine learning solution for body pose tracking and estimation. The classifier obtained an accuracy score of 92.92%, precision scores of no lower

Figure 4: Application Prototype

than 81%, recall scores of no lower than 83% and f1-scores of no lower than 86% across all 6 squat posture classes plus the newly introduced standing posture class on the test set. Although the classifier produced promising results, important limitations which affects the classifier's practical use in the real world were identified and need to be considered. These include the contextual assumptions and system requirements needed to be met to ensure the model to properly run and make classifications, and the small sample from which the data used for this project was gathered.

The researcher recommends the following points of improvement for further studies on this topic:

1. The researcher decided to use the Support Vector Machine as the type of learning model for the classifier to minimize the computational resources needed to make classifications. However, this resulted in the learning model to only be able to classify postures. Newer types of learning models especially those which consider the temporal context of the data may also allow for the classification of squat repetitions, or perhaps even an entire squat set of repetitions.

2. The researcher was only able to gather data from 10 individuals due to time constraints. A larger sample from which the data is gathered is highly likely to improve the robustness of the model.

6. REFERENCES

[1] N. Hongu, M. J. Wells, P. J. Gallaway and P. Bilgic, Resistance Training: Health Benefits and Recommendations, The University of Arizona, College of Agriculture and Life Sciences, Tucson, Arizona, Apr 2015.

[2] P. J. O'Connor, M. P. Herring and A. Caravalho, Mental Health Benefits of Strength Training in Adults, *American Journal of Lifestyle Medicine*, Sept 2010, pp. 377-396.

[3] B. S. Shaw, I. Shaw and G. A. Brown, Resistance exercise is medicine: Strength training in health promotion and rehabilitation, *International Journal of Therapy and Rehabilitation*, Aug 2015, pp. 385-389.

[4] C. S. Jones, C. Christensen and M. Young, Weight Training Injury Trends, *The Physician and Sports Medicine*, Jul 2000.

[5] C. Muonwe, S. C. Nwobi, C. J. Alumona, C. Okeke, C. A. Nwanne, I. B. Nwosu, I. A. Amaechi, C. A. Okonkwo, O. Umeonwuka, E. C. Ekediegwu, U. N. Akobundu and S. M. Maduagwu, Prevalence and Pattern of Musculoskeletal Injuries among Recreational Weightlifters in Nnewi, Nigeria, *International Journal of Sports and Exercise Medicine*, 2021.

[6] A. M. Myers, N. W. Beam and J. D. Fakhoury, Resistance training for children and adolescents, *Translational Pediatrics*, Mar 2017, pp. 137-143.

[7] P. R. Stricker, A. D. Faigenbaum, T. M. McCambridge and AAP Council on Sports Medicine and Fitness, Resistance Training for Children and Adolescents, *Pediatrics*, Jun 2020.

[8] F. A. Barwais, Physical Activity at Home During the COVID-19 Pandemic in the Two Most-affected Cities in Saudi Arabia, *The Open Public Health Journal*, Jul 2020, pp. 470-476.

[9] H. Kaur, T. Singh, Y. K. Arya and S. Mittal, Physical Fitness and Exercise During the COVID-19 Pandemic: A Qualitative Enquiry, *Frontiers in Psychology*, Oct 2020.

[10] J. A. Petersen, C. Naish, D. Ghoneim, J. L. Cabaj, P. K. Doyle-Baker, G. R. McCormack, Impact of the COVID-19 Pandemic on Physical Activity and Sedentary Behavior: A Qualitative Study in a Canadian City, *International Journal of Environmental Research and Public Health*, Apr 2021.

[11] C. Freytag, How To Do A Basic Squat, `https://gethealthyu.com/exercise/basic-squat/` (Accessed: April 25, 2022).

[12] A. Depari, P. Ferrari, A. Flammini, S. Rinaldi and E. Sisinni, Lightweight Machine Learning-Based Approach for Supervision of Fitness Workout, *2019 IEEE Sensors Applications Symposium (SAS)*, 2019, pp. 1-6.

[13] J. Lee, H. Joo, J. Lee and Y. Chee, Automatic Classification of Squat Posture Using Inertial Sensors: Deep Learning Approach, *Sensors*, Jan 2020.

[14] R. Ogata, E. Simo-Serra, S. IIizuka and H. Ishikawa, Temporal Distance Matrices for Squat Classification, *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, pp. 2533-2542.

[15] G. Park, Design of Mobile Personal Workout Assistant Using Deep Learning, Seoul National University, Aug 2020.

[16] MediaPipe, MediaPipe Pose, `https://google.github.io/mediapipe/solutions/pose.html` (retrieved February 18, 2022).