

Smart Coaching: Enhancing Weightlifting and Preventing Injuries

Ammar Yasser¹, Doha Tariq²,
Radwa Samy³, Mennat Allah Hassan⁴
Misr International University
Computer Science

Ayman Atia⁵
HCI-LAB, Faculty of Computers,
Helwan University,
Misr International University
Computer Science

Abstract—Getting injured is one of the most devastating and dangerous challenges that an athlete can go through and if it is a big injury it could end his/her athletic career. In this paper, we propose a system to automate the idea of coaching an athlete, by using an IR Camera (Microsoft Kinect Xbox 360) to detect the misplaced joints of the athlete while doing the lift, and alerting the athlete before an injury can occur. We are now able to detect if the lift was correct or wrong and to detect what kind of mistake has been done in the lift by the athlete by using the Fast Dynamic time warping (FastDTW) method. The FastDTW method has outperformed other classification methods and can achieve recognition with 100% accuracy for dependent user movements.

Keywords—Weightlifting; joints; KNN; fastDTW; Naive Bayes; SVM; detecting injuries; machine learning; IR camera

I. INTRODUCTION

Weight training is one of the most injuries affecting training methods due to wrong techniques of lifting the weights. Although few systems were developed to analyze some weightlifting exercises, they still need a coach's view to correct the wrong techniques. Weightlifting is the sport of lifting a heavy barbell in various different ways, and it has been an Olympic sport since 1922. The main lifts of Olympic weightlifting are the Snatch and the Clean&Jerk. Those lifts are called complex lifts that consist of several movements attached together to form one whole lift.

Three of those fundamental movements are the Deadlift, the Squat and the Shoulder Press. In the 'Deadlift' as shown in Fig. 1 the athlete just picks up the weighted barbell from the ground. In the 'Squat' as shown in Fig. 2 the athlete sits down and stands up with the barbell on his/her back. The 'Shoulder Press' as shown in Fig. 3 is pressing the barbell from his/her shoulder to overhead.

Each movement has a special pattern that must be done with at least 90% of the Right technique or an injury would occur otherwise [9]. In the Deadlift, the spine has to be straight in one line as shown in Fig. 1(a), if it is rounded as Fig. 1(b), the chance of injuring the lower back is high. In the Squat, back should be straight, knees behind the toes and heels on the floor as shown in Fig. 2(a), if the back is not straight and knees surpass the toes as shown in Fig. 2(b), there is a high risk of injuring the knee joints. In the Shoulder press, the bend in the back must be as limited as possible, if it is too bent as shown in Fig. 3(b), an injury will occur in the back.



Fig. 1. Deadlift: a) Right. b) Wrong.



Fig. 2. Squat: a) Right. b) Wrong.

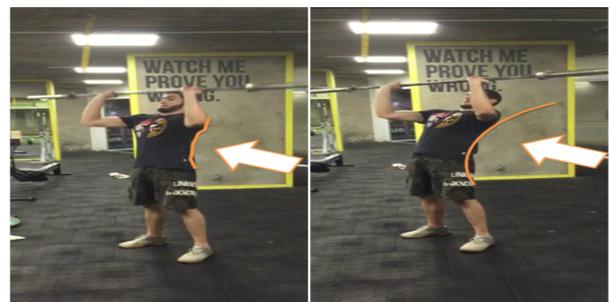


Fig. 3. Shoulder Press: a) Right. b) Wrong.

Teaching the right techniques of these movements is a task that requires an expert at lifting weights. We present a system that has the ability to coach an athlete at any level to do those movements injury-free.

Preventing injuries and showing the right technique of the

lift remains our core focus. Working on the fundamental lifts in weightlifting can assure right movement patterns and injury-free lifting the two complex lifts that consist of those three lifts we mentioned before.

For example, Fig. 1 shows the difference between the wrong and the right technique of the 'Deadlift'; on the left image, it shows the straight spine while on the right one it shows the bent spine which causes an injury in the back. Most injuries that occur during wrong techniques are joint injuries because the joint is the main connection between all bones. It is vital to detect how the joints move to correct the movement pattern. We utilize the infrared camera for detecting the coordinates of the joints of the athlete while doing the lift, which is then tested and classified according to our training data set.

The training data set consists of joint coordinates extracted from the infrared camera videos taken of coaches and athletes doing the three movements with the right and the wrong techniques.

The infrared camera we are using (IR Camera) reading coordinates are fixed, so if the camera's placement changed from one place to the other, the joint coordinates will not be as accurate. This problem had solved by using skeletal tracking, which is a fixed place that the athlete has to start from, so the joint readings are more accurate.

Several classifiers were tested such as Naive Bayes, KNN and Fast DTW. Classification of such a data set faces many challenges, such as the different number of frames from one video to the other, one athlete may take 3 seconds to complete the full lift and one may take 5 seconds. This problem had solved by employing the FastDTW classifier, as it ignores the speed of the video and classifies according to the sequence of frames.

II. RELATED WORK

Many kinds of research and projects used similar algorithms and equipment in various scopes. Each of them reached different accuracy and numerous results.

A. Coaching

In the past, some research has used a diversion of methods to coach people to improve their performances doing their daily activities.

Xin Jin et al. [7], presented a system based on visual guidance that helps users perform exercises with the right techniques. The system has consist of two phases, the User phase and the Dataset phase which is the analysis phase. Their system used the DTW algorithm to compare the user's techniques to the already stored videos in the database to measure the accuracy. Then it guides the user to fix their technique to meet the standard technique. They used the Kinect to get the positions of the joints of the user's body. They managed to improve the accuracy of the users' performance by 72.98%.

Pradeep Kumar et al. [10], proposed a real-time virtual trainer. That used the Random Forest Classifier to recognize the exercises performed by the users. They prepared their

dataset by choosing 5 ideal exercises that can be performed by anyone in daily life, these five exercises were done by 3 different Fitness trainers. Their experiment's outcome was 96% accurate.

Edwin W Trejo et al. [16], presents a system that uses the Ada-Boost algorithm to analyze 10 clips of each of the 6 yoga poses in the dataset. An avatar of the user will change colors according to the rightness of the pose. 94.78% is the accuracy level they reached training people to perfect their yoga poses.

Hua-Tsung Chen et al. [4], used Contour and Skeleton Computation to capture binary maps of the body. Then used Feature Axis Extraction algorithm to extract certain points on the body to measure the correctness of the three Yoga poses. To then assist yogis to improve their postures. They reached an overall accuracy of 98.67%.

Out of all the previous researches, our system has some concepts in common with them and differs in others. For instance (i) The use of the Kinect in virtual training. (ii) We used the DTW algorithm. (iii) We target workout and fitness exercises.

B. Weightlifting

Many applications and projects recently appeared to help their users with their workout but very few can guide the user with their Weightlifting techniques.

Pichamon Srisen et al. [15], worked on detecting the 20 main joints in the human body that are involved in weightlifting. To do that they used the infrared camera (Kinect) and the Lucas-Kanade optical flow algorithm. Their accuracy reached 80.5% for hands, feet and knees.

Anargyros Chatzitofis et al. [3], created a weightlifting electronic assistant. Using the infrared camera (Kinect) to detect the human structure. They calculated: i) The Weightlifting Bar Position. ii) The Weightlifting Bar Angle. iii) The Weightlifting Bar Velocity. iv) The Knee Angle Calculation.

Abdul Monem S Rahma et al. [13], proposed an analysis system that monitors weightlifters performing the Snatch and Clean&Jerk moves. Using the Correlation algorithm they managed to get to 75% accuracy.

Perfecting the accuracy of the lift, minimizing the chances of injury and facilitating the training techniques are things we have worked to improve in our system.

C. Camera based Techniques

In trying to make the electronic training or coaching concept more approachable for works, the human started to add the virtual aspect using various virtual aiding technologies.

Orasa Patsadu et al. [12], used an infrared camera (Kinect) to recognized different gesture patterns. they started to test on an individual video for different six human bodies and gender. They used a different data mining classification algorithms to recognize those gestures parent in their videos and to classify them, such as Back-Propagation Neural Network (BPNN), Support Vector Machine (SVM), Decision Tree, and Naive Bayes. Only two of those algorithms showed the highest accuracy. These two algorithms are the SVM which showed 99.75% accuracy and the Naive Bayes showed 81.94% accuracy.

Frederik Wieh et al. [17], used an infrared camera (Kinect 2) to recognize the skeleton of the climber doing a successful ascent.

Sai Prakash Reddy Gaddam et al. [6], used an infrared camera (Kinect 2), Vicon Cameras and Force Plates to compare the force the human's Jumps. The results collected from the Kinect were close to those collected from the Force Plates.

Sean Clarkson et al. [5], used four infrared cameras (Kinect) to mount their test objects. They used those cameras to take shots of four different sized cylinders in different positions multiple times. To compare between the output and the ISO standards. The third and fourth cylinders met the ISO 20685-1 requirement standard with large girth and confidence of 95%.

As part of our system, we used one infrared camera (Kinect) to capture the athlete performing the lift. To cluster and classify them as those lifts were there are the right techniques or detect as the wrong techniques which may cause injuries for the athletes.

D. Classifiers

In order to run tests on the visual data collected from the IR Camera and other visual sensors, some research started to use a wide number of classifiers such as KNN, SVM, GMM, DTW, etc.

Alina Delia Calin et al. [2], proposes a system that uses multiple IR cameras as (Kinect 1 and Kinect 2) to capture poses and gestures. The system tested 41 classifiers such as Simple Logistic, Multilayer Perceptron, Random Forest and Naive Bayes. Then combined all obtained results from the Kinect 2 data sets and the ones collected from Kinect 1 to compare classifiers' performances. She compared the time, accuracy and precision to build the models. Some of the highest accuracy results she reached are for the Multilayer Perceptron algorithm scored accuracy: 99.08%, precision: 99.1%, Random Forest scored accuracy 98.957%, precision 99%.

Sowmya Kasturi et al. [8], used the support vector machine (SVM) to detect and classify which action was a fall and which was done deliberately. Their method showed a total training accuracy of 99.7% and a total testing accuracy of 96.3%.

Lichao Zhan et al. [19], they presented the development of a single-user adaptive scoring system for Golf Swing by using an infrared camera (Kinect). They used the support vector machine (SVM) and Gaussian Mixture Model (GMM) to classify and cluster the swing. Their system improved the accuracy of Golf Swing recognition by 84.1%.

Manus Ross et al. [14], used an IR Camera (Kinect) and the RGB-D sensor to detect, monitor, count and record student gestures, postures, facial expressions, and verbalizations in order to produce data for determining student attentiveness. The data collected is clustered into two clusters using the K-means algorithm. The SVM was then used to classify the clustered data to establish decision boundaries.

Yi-Hua Zhou et al. [20], used the support vector machine (SVM) classifier to classify the shot type from a football video. They also used GMM to remove the grass from the video and HSV for the color distribution. For the edge distribution, they

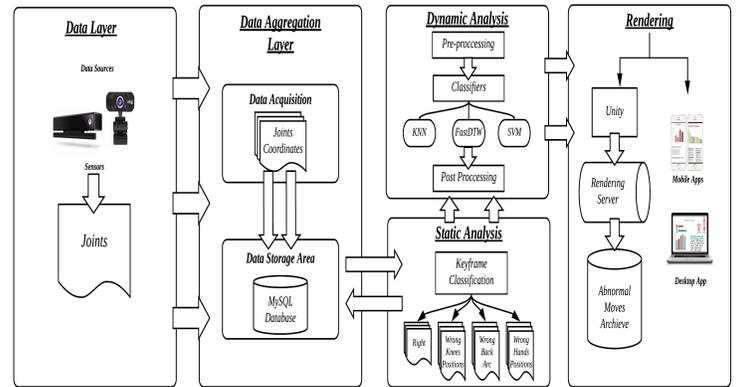


Fig. 4. The System's Block Diagram.

used the canny operator where they proposed that had high detection performance. The average precision that they attain is 92%.

Choubik Youness et al. [18], used an IR Camera (Kinect) to detect the human skeleton to classify the human poses. They used the support vector machine (SVM), artificial neural networks (ANN), k-nearest neighbors (KNN) and Bayes classifier (BC). They tested 100 examples for each pose. They reach the accuracy to 100% in each classifier but with different amount of training data that they tested. For the support vector machine (SVM), the Linear Kernel needed 44%, the Polynomial Kernel needed 66% and the RBF Kernel needed 55%. For the artificial neural networks (ANN), they used for the Sigmoid and Gaussian function needed 44%. For the k-nearest neighbors (KNN), they needed 66%. But for the Bayes classifier (BC) it was the high accuracy reach is 99.9% at 88% of the train data.

In our system, we have tested some of the previously mentioned algorithms such as the SVM, FastDTW, Naive Bayes, KNN and the Random Forest. Different algorithms have given various results.

III. METHODOLOGY

A. Pre-Processing

First, the system used the IR camera and the Camera to capture the Lift. The IR Camera detected 20 Joint from the human body as shown in Fig. 5. These joints are put in front of the athlete on a screen. We worked on Shoulder Press, Deadlift and Squat style. For Shoulder press, we detected six joints in the athlete's body which are Wrist Left, Wrist Right, Shoulder Left, Shoulder Right, Center Shoulder, and Spine as shown in Fig. 5. For Deadlift and Squat, we detected six joints in the athlete's body which are Knee Left, Knee Right, Shoulder Left, Shoulder Right, Center Shoulder, and Spine as shown in Fig. 5.

The IR Camera was used to get the X, Y, Z coordinates of each joint, that has been detected for each movement. And the

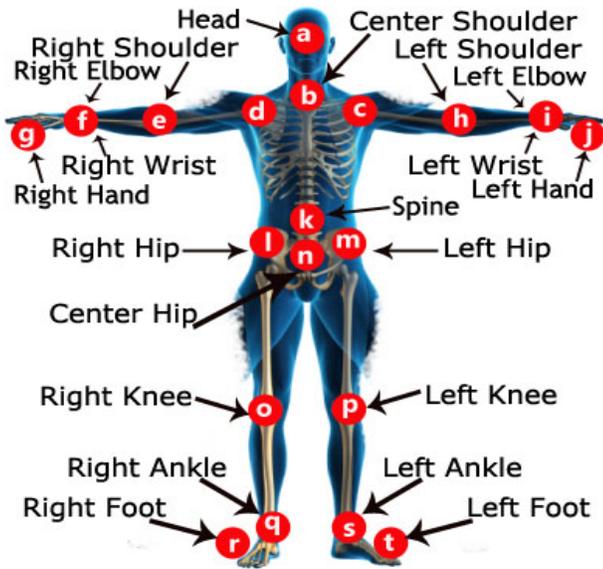


Fig. 5. The Joints that detected by the IR Camera.

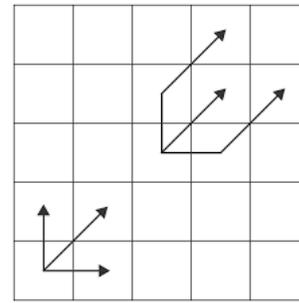


Fig. 6. Creating Cost Matrix.

Camera recorded the whole lift. It was put on either side of the athlete. Then, the data received from the IR Camera and the Camera were stored in our data storage as shown in the block diagram (Fig. 4).

B. Classifiers

In this phase, the system used different classifiers to detect the result of the last lift. FastDTW was used to classify the lift as right or wrong. The System then allows the coach to choose the most accurate of the recently classified lift to be added to the template data from the data storage.

1) *The k-nearest neighbors(KNN)*: The k-nearest neighbors are one of the oldest and simplest methods of classification. The idea behind the KNN algorithm is quite simple. Given the video of the last lift to get the coordinates of each joint that has been detected for each movement.

And set of all the videos that have the coordinates of each joint in our data set and their labeled neighbors from the data storage in the template table. The task of the classifier is to predict the video's class label based on the class labels in the set. Using the 5 nearest neighbors using the majority vote by the Euclidean Distance as shown in equation (1) [1].

$$d(p, q) = d(q, p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (1)$$

p: points of tested video, q: points of each video from dataset.

2) *Fast Dynamic Time Warping (FastDTW)*: Time series analysis, dynamic time warping (DTW) is one of the most used algorithms for measuring the similarity between two temporal

sequences for the same action but different in time and speed. It has been designed especially for time series analysis, because it helps to ignore Shifts in the time dimension, ignore Speeds of two-time series.

First, the FastDTW start to create a cost matrix between the test lift's coordinates of the athlete body and each video in the data set.

$$Dist(I, J) = Dist(I, J) + \min \begin{cases} Dist(I+1, J) \\ Dist(I, J+1) \\ Dist(I+1, J+1) \end{cases} \quad (2)$$

I: tested lift points, J: one lifts points from data set

FastDTW starts to create each point in the matrix that between the test lift points and each lift points in the data set, by getting the minimum value between the two points and the minimum of its neighbors as shown in Fig. 6.

Second, the FastDTW uses backtracking & greedy search in the cost matrix to get the distance between the two lifts.

$$Dist(W) = \sum_{K=1}^{K=L} Dist(W_{KI}, W_{KJ}) \quad (3)$$

L: Last left point in the cost matrix, W: the cost matrix, K: each cell in the matrix

FastDTW starts to get the distance between the two lifts, by adding the point from the top left of the cost matrix cell. Then go to get the minimum of its neighbors until it reaches the first cell of the matrix.

In the end, after getting the distance between each lift in the data set and the tested video. It starts to get the minimum one of them to get the label of the lift from the data set to be labeled the lift as it been classified.

3) *Support Vector Machine (SVM)*: Support Vector Machines (SVM) uses supervised learning data. Its basic idea is to find an optimal hyperplane margin that properly separates data by choosing a vector from each set. Keeping in consideration that has to be the furthest away as possible from all data. SVM categorizes each group of videos by their labels grouping them

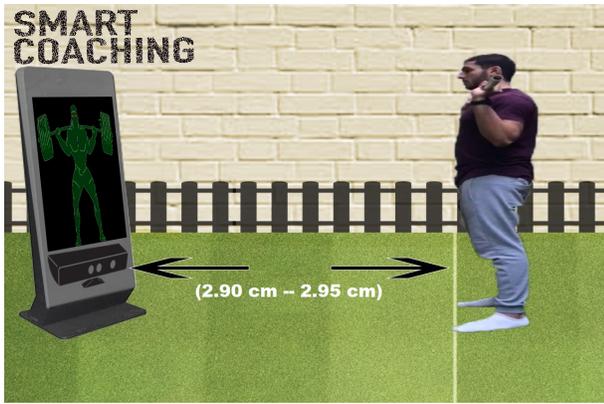


Fig. 7. The System overview.

together. It then starts calculating the new video's distance to define which class it belongs to.

$$\begin{aligned}
 F &= [J_{1x}, J_{1y}, J_{1z}, J_{2x}, J_{2y}, J_{2z}, J_{3x}, J_{3y}, J_{3z}]. \\
 L &= [F_1, F_2, F_3, \dots, F_n]. \\
 R &= \sqrt{(qF_1 - pF_1)^2 + (qF_2 - pF_2)^2 + \dots + (qF_n - pF_n)^2} \quad (4)
 \end{aligned}$$

J: Joint, F: Frame, L: Lift, R: Result, q: Test Lift & p: from data set.

So, our vector that we used in SVM is the X, Y, Z coordinates of each joint in each frame of the video. Then, it starts to compare each position of each joint in each frame in the lift and each video in the data set. In the end, it seems the nearest video labels to get the result.

C. Post-Processing

Finally, our system lets coach choose which of the newly classified videos is accurate enough to be added in the template table. To start building a machine learning sequence that allows the system to improve the results.

IV. EXPERIMENTAL SETUP

The IR Camera was placed on a box with height, 47 cm and the athlete is standing away from it with a distance between (2.90cm -2.95 cm) as shown in Fig. 7. There, were 10 athletes, all were males, aged between 20 to 30 doing the Shoulder Press movement, deadlift, and squad. The aim of the experiment is to test different algorithms for classification of movements (user dependent and user independent), respectively. Our experiment has been done in Professional Gym.

A. User Dependent Experiment

A senior coach has been asked to be a subject for our experiment, and he had to build his own training data set, he has to record the lift 10 times per each class. After that, we have asked 1 athlete to test all the three movements for nine times. here are the results are shown in Table I.

TABLE I. RESULTS OF THE USER DEPENDENT EXPERIMENT ON THE THREE MOVEMENTS

Movement	FastDTW	KNN	SVM	Naive Bayes
Shoulder Press	100%	77.78%	88.89%	66.67%
Deadlift	100%	100%	100%	88.89%
Squats	100%	66.67%	77.78%	66.67%

TABLE II. RESULTS OF THE USER INDEPENDENT EXPERIMENT ON THE THREE MOVEMENTS

Movement	FastDTW	KNN	SVM	Naive Bayes
Shoulder Press	91%	86%	70%	68%
Deadlift	100%	100%	93%	87%
Squats	98%	60%	40%	56%

B. User Independent Experiment

A senior coach and two beginner athletes have been asked to be a subject for our second experiment, they have to record their lift 15 times per each movement. then we start classifier each lift by using different algorithms and here is the results shown in Table II.

FastDTW has shown the best accuracy compared to KNN, SVM and Naive base with an accuracy of 100%. FastDTW shows the highest accuracy because it almost works with different time series analysis to measuring the similarity between two temporal sequences for the same action but different in time and speed. The naive base has shown the worst accuracy, based on Lindsay et al. [11] presented that Naive base algorithm shows low accuracy with time series analysis, due to Naive Bayes learner invalidly assumes independence of attributes of the lift.

V. CONCLUSION AND FUTURE WORK

We present a system that automates the process of coaching an athlete through the fundamental lifts which are: the Squat, the Deadlift and the Shoulder Press. Decreasing the injuries that occur due to these three movements is the main focus. In future work, the scope of the system could grow by adding more movements in weightlifting sport. Our future work, to add more different movement in weightlifting, and the wrong moves that bring different injuries for the athletes. Also to normalize the dataset to avoid the 2.90cm range that has been used in this paper.

ACKNOWLEDGMENT

We would like to express our gratitude and appreciation to Dr. Ayman Atia the head of our research group in the department of Computer Science at MIU. He always provides us with promising research directions and continuous guidance.

REFERENCES

- [1] DA Adeniyi, Z Wei, and Y Yongquan. Automated web usage data mining and recommendation system using k-nearest neighbor (knn) classification method. *Applied Computing and Informatics*, 12(1):90-108, 2016.

- [2] Alina Delia Călin. Variation of pose and gesture recognition accuracy using two kinect versions. In *2016 International Symposium on Innovations in Intelligent Systems and Applications (INISTA)*, pages 1–7. IEEE, 2016.
- [3] Anargyros Chatzitofis, Nicholas Vretos, Dimitrios Zarpalas, and Petros Daras. Three-dimensional monitoring of weightlifting for computer assisted training. In *Proceedings of the virtual reality international conference: Laval virtual*, page 3. ACM, 2013.
- [4] Hua-Tsung Chen, Yu-Zhen He, Chien-Li Chou, Suh-Yin Lee, Bao-Shuh P Lin, and Jen-Yu Yu. Computer-assisted self-training system for sports exercise using kinects. In *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pages 1–4. IEEE, 2013.
- [5] Sean Clarkson, Jon Wheat, Ben Heller, and Simon Choppin. Assessment of a microsoft kinect-based 3d scanning system for taking body segment girth measurements: a comparison to isak and iso standards. *Journal of sports sciences*, 34(11):1006–1014, 2016.
- [6] Sai Prakash Reddy Gaddam, Mukesh Kumar Chippa, Shivakumar Sasstry, Anthony Ange, Visar Berki, and Brian L Davis. Estimating forces during exercise activity using non-invasive kinect camera. In *2015 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 825–828. IEEE, 2015.
- [7] Xin Jin, Yuan Yao, Qiliang Jiang, Xingying Huang, Jianyi Zhang, Xiaokun Zhang, and Kejun Zhang. Virtual personal trainer via the kinect sensor. In *Communication Technology (ICCT), 2015 IEEE 16th International Conference on*, pages 460–463. IEEE, 2015.
- [8] Sowmya Kasturi and Kang-Hyun Jo. Classification of human fall in top viewed kinect depth images using binary support vector machine. In *Human System Interactions (HSI), 2017 10th International Conference on*, pages 144–147. IEEE, 2017.
- [9] Dr. Stephen Kirk. Common injuries from weight lifting. <http://www.utahorthopediccenters.com/common-injuries-from-weight-lifting/>, 2018.
- [10] Pradeep Kumar, Rajkumar Saini, Mahendra Yadava, Partha Pratim Roy, Debi Prosad Dogra, and Raman Balasubramanian. Virtual trainer with real-time feedback using kinect sensor. In *IEEE Region 10 Symposium (TENSymp), 2017*, pages 1–5. IEEE, 2017.
- [11] David Lindsay and Siân Cox. Effective probability forecasting for time series data using standard machine learning techniques. In *International Conference on Pattern Recognition and Image Analysis*, pages 35–44. Springer, 2005.
- [12] Orasa Patsadu, Chakarida Nukoolkit, and Bunthit Watanapa. Human gesture recognition using kinect camera. In *2012 Ninth International Conference on Computer Science and Software Engineering (JCSSE)*, pages 28–32. IEEE, 2012.
- [13] Abdul Monem S Rahma et al. Proposing an analysis system to monitoring weightlifting based on training (snatch and clean and jerk). *Baghdad Science Journal*, 15(4):493–502, 2018.
- [14] Manus Ross, Corey A Graves, John W Campbell, and Jung H Kim. Using support vector machines to classify student attentiveness for the development of personalized learning systems. In *Machine Learning and Applications (ICMLA), 2013 12th International Conference on*, volume 1, pages 325–328. IEEE, 2013.
- [15] Pichamon Srisen, Sansanee Auephanwiriyaikul, Nipon Theera-Umpon, and Samatchai Chamnongkich. Kinect joints correction using optical flow for weightlifting videos. In *Computational Intelligence, Modelling and Simulation (CIMSIM), 2015 Seventh International Conference on*, pages 37–42. IEEE, 2015.
- [16] Edwin W Trejo and Peiji Jiang Yuan. Recognition of yoga poses through an interactive system with kinect device. In *2018 2nd International Conference on Robotics and Automation Sciences (ICRAS)*, pages 1–5. IEEE, 2018.
- [17] Frederik Wiehr, Felix Kosmalla, Florian Daiber, and Antonio Krüger. betacube: Enhancing training for climbing by a self-calibrating camera-projection unit. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1998–2004. ACM, 2016.
- [18] Choubik Youness and Mahmoudi Abdelhak. Machine learning for real time poses classification using kinect skeleton data. In *2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV)*, pages 307–311. IEEE, 2016.
- [19] Lichao Zhang, Jui-Chien Hsieh, Tsu-Te Ting, Yi-Chi Huang, Ya-Chih Ho, and Lin-Kai Ku. A kinect based golf swing score and grade system using gmm and svm. In *Image and Signal Processing (CISP), 2012 5th International Congress on*, pages 711–715. IEEE, 2012.
- [20] Yi-Hua Zhou, Yuan-Da Cao, Long-Fei Zhang, and Hong-Xin Zhang. An svm-based soccer video shot classification. In *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*, volume 9, pages 5398–5403. IEEE, 2005.