BowlingDL: A Deep Learning-Based Bowling Players Pose Estimation and Classification

2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC) | 978-1-6654-7275-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAISC56366.2023.10085434

Nourah Fahad Janbi
Department of Information Technology, College of Computing and Information Technology at Khulais
University of Jeddah
Jeddah, Saudi Arabia
nfjanbi@uj.edu.sa

Nada Almuaythir
Services and Electronic Solutions Sector
Tatweer Company
Riyadh, Saudi Arabia
Nada.Almuaythir@gmail.com

***Abstract*—Human Pose Estimation (HPE) is one of the trending areas of research among artificial intelligent research. It has gained a lot of attention due to its versatile potential applications in various domains including transportation, healthcare, gaming, augmented reality, and sports. HPE can be used to build sports analytics, personalized training, and self-learning systems which allow players, athletes, and trainers to improve the training quality by evaluating various human poses detected from images or videos. As far as we know none of the exciting works considered developing a pose estimation and classification framework for bowling players. Therefore, in this paper, we proposed a deep-learning approach for bowling players' pose estimation and classification. It uses our proposed Bowling Deep-Learning (BowlingDL) model along with the MoveNet model for bowling players' pose estimation and classification. The MoveNet model detects various key points in human pose and the BowlingDL model classifies the detected bowling player's poses into five different classes. For model training and evaluation, we collected and labelled our own dataset as no dataset was found for bowling posture. Our proposed model achieved 80% accuracy for the training dataset and 83% accuracy for the testing dataset. In addition, a smart mobile application for bowling players was developed where an edge-friendly version of BowlingDL—generated using TensorFlow Lite—was deployed.**

Keywords—Artificial Intelligent, Edge Intelligence, Mobile Intelligence, Deep Learning, Convolutional Neural Networks, Human Pose Estimation, Pose Classification, Smart Applications, Sports Analytics.

I. Introduction

Human Pose Estimation (HPE) is one of the computer vision problems where human body parts are located. It aims to understand the human body's geometry and motion. HPE has two basic steps: collecting body biomechanical characteristics and parameters from an image or series of images (video), and identifying or analyzing a human pose/s using the collected data. Deep learning is one of the widely used methods for pose identification and classification wherein the model takes body pose parameters as an input and predicts a posing class.

The continuous advancement in software and hardware technologies has significantly boosted the development and implementation of HPE in recent years [1]. Moreover, HPE and pose classification have gained a lot of the researchers' attention due to their versatile potential applications including video surveillance [2], daily activities recognition [3], assisted living [4], driver assistant systems [5], sign language detection [6], gaming and character animation [7], and sports monitoring and training [8].

One of the interesting application domains of HPE is the sport and physical activity domain where HPE can be used to build applications such as sports analytics [9], personalized training [10], and self-learning systems [11]. These applications aim to assist athletes, trainers, other sports experts, and anybody trying to practice sports or physical activities in learning, evaluating, and improving their performance.

Looking at existing works of HPE in sports, it is believed that the applicability of HPE in sports is at its early stages and several sports applications need to be developed and tested by the research and developer community [12]. Existing works have focused on sports such as Soccer and Football [13, 14], Tennis [15], Fitness [10, 11], Yoga [16, 17], etc. However, as far as we know, none of the exciting works considered HPE for bowling players. Therefore, the novelty of this work lies in the development and evaluation of a pose estimation and classification framework for bowling players specifically and the main contributions can be outlined as follows:

- This is the first paper where a deep-learning approach for bowling players' pose estimation and classification is introduced, developed, and evaluated.
- Collecting and labelling the first Bowling Posture Dataset consisting of 5 different classes of bowling posture.
- Developing, training, and evaluating the Bowling Deep-Learning (BowlingDL) model for bowling players' pose classification.
- Developing and deploying a smart mobile application for bowling players where an edge-friendly version of the BowlingDL and the MoveNet models are generated and deployed for bowling players' pose estimation and classification.

The rest of the paper is organized as follows: Section 2 reviews the related works. Section 3 explains our methodology. Section 4 describes our proposed BowlingDL framework for pose estimation and classification. Section 5 concludes the paper.

II. Related Works

HPE is a popular topic in the computer vision field, researchers have proposed different approaches, models, and architectures for different applications domains including video surveillance [2], daily activities recognition [3], assisted living [4], driver assistant systems [5], sign language detection [6], gaming and character animation [7], and sports monitoring and education [8]. Cormier et al. [1] provided a detailed survey on the progress of HPE research.

In the sports field, existing research has focused on sports such as Soccer and Football [13,14] Tennis [15], Fitness [10, 11], Yoga [16, 17], etc. Table 1 provides a summary of some of the related works. For example, Arbues-Sanguesa et al. [13] used the OpenPose model with contextual information to estimate soccer players' body orientation. Sypetkowski et al. [14] focused on estimating football player poses in very low-resolution images affected by weather conditions, shadows, or reflections. Shimizu et al. [15] developed a tennis player future shot direction estimation method based on HPE and player position. Garg et al. [16] proposed an architecture where skeletonized images generated using the MediaPipe model are fed to a Convolutional Neural Network (CNN) model to classify yoga postures. Kothari [17] adopted a hybrid CNN and recurrent neural network (RNN) model for yoga pose classification.

In the fitness domain, many researchers have focused on training assistance systems. For instance, Wang et al. [10] developed a personalized athletic training assistance system including human tracking, pose estimation, pose correction, and visual suggestion. Similarly, Zou et al. [11] presented a fitness training hardware and software design with both fitness training and motion correction based on HPE. A full review of HPE for training assistance systems can be found in [18].

Table 1. Related work summary

Ref.	Application	Approach
Arbues-Sanguesa et al. [13]	Soccer player Orientation	OpenPose with a super-resolution network
Sypetkowski et al. [14]	Football Players Pose Estimation	Cascaded Pyramid Networks (CPN) model
Wang et al. [10]	Training Assistance	Tracking, HPE, correction, and suggestion
Zou et al. [11]	Intelligent Fitness Trainer	HPE and pose correction
Shimizu et al. [15]	Tennis Future Shot Direction	HPE and player position
Garg et al. [16]	Yoga Pose Estimation	MediaPipe
Kothari [17]	Yoga Pose Estimation	OpenPose, RNN, and CNN

Looking at recent works, literature reviews and surveys on HPE [1,18,19] and HPE for sports and physical activities specifically [12], it is noticed that research on HPE in sports and physical activities is still in its early stages and several sports applications need to be developed and tested by the research and developer community. Moreover, as far as we know, none of the exciting works considered HPE for bowling players. Therefore, in this work, we developed and evaluated a pose estimation and classification framework for bowling players.

III. Methodology

Fig. 1 captures the methodology that this research has followed. The first step of our pipeline is data collection where the bowling posture dataset is collected and labelled. In the preprocessing stage, for each image in the bowling posture dataset, key points of the human pose are estimated. After that, the BowlingDL model is designed, and the generated key point dataset is used for model training and evaluation. Model design, training, and evaluation were iterative until a reasonable result was achieved. The final BowlingDL model is saved in a proper file format and deployed in an AI Bowling Application (in our case, it was an Android application).

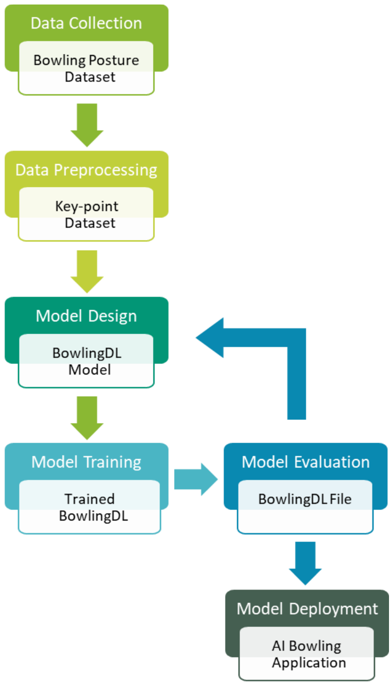

Fig. 1. Methodology

IV. Bowling Deep-Learning (BowlingDL) Pose Estimation and Classification

This section provides a detailed description of our BowlingDL framework for pose estimation and classification. First, the Bowling Posture Dataset is described (Section A). Section B discusses data preprocessing and Section C provides detailed design architecture of the BowlingDL model. Finally, Section D discusses model training and evaluation, and Section E describes our deployment and testing stage including the developed application.

A. Bowling Posture Dataset

In this research, we built our dataset from scratch, as no image dataset was found for bowling posture. We collected our dataset by capturing images from videos available on YouTube [20]. These include videos from world bowling associations and competitions (e.g. The Professional Bowlers Association [21]), private academies (e.g. The National Bowling Academy [22] and Inside Bowling [23]), private personal channels specialized in bowling (e.g. The House Bowling [24]), and many daily life bloggers. The selected images capture only the posture of the bowler just before he/she throws the ball. A total of 193 images were collected and labelled into 5 categories: one-handed left, one-handed right, two-handed left, two-handed right, and improper. Table

2 lists the number of images for each category in the Bowling Posture Dataset.

TABLE 2. BOWLING POSTURE DATASET CHARACTERISTICS

Class	**# of Images**
One-Handed Left	29
One-Handed Right	32
Two-Handed Left	29
Two-Handed Right	40
Improper	63
Total	**193**

B. Data Preprocessing

In the data preprocessing stage, images are fed to the MoveNet [25] model which is used for the human pose estimation phase. The MoveNet model is one of the models provided by TensorFlow [26] for human pose estimation. It has two variants Lightning and Thunder, in this work we used the MoveNet Thunder model for both the training and evaluation phases. The model detects 17 key points of the human body including the nose, eyes, ears, shoulders, elbows, wrists, hips, knees, and ankles. For each image, the model returns an array with the shape [1, 1, 17, 3] which comprises x and y coordinates and scores for each key point. Fig. 2 captures an example of key points detected by the MoveNet model. For training and evaluation, all images on the Bowling Posture dataset are preprocessed and key points information is stored to be used for BowlingDL training.

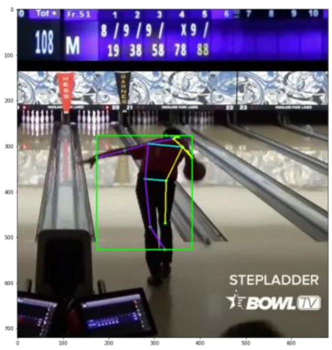

Fig. 2. MoveNet bowler keypoint detection

C. Bowling Deep Learning (BowlingDL) Model Design

As Convolutional Neural Networks (CNN) has approved their strength in image classification and recognition problems, we designed our BowlingDL model based on CNN. The BowlingDL model takes the key point information of bowling player pose, which is detected by the MoveNet model, and classifies it into one of the 5 categories: one-handed left, one-handed right, two-handed left, two-handed right, and improper. The model was implemented using the Keras library in Python. Fig. 3 captures the BowlingDL model architecture. It consists of multiple layers including Dropout, Flatten, TFOpLambd, and Dense layers and utilizes different activation functions including ReLU6 and SoftMax.

In the architecture, first, the input landmarks are converted into a pose embedding by reshaping the flat input into (17,3) shape representing the 17 key points and x, y, and score values for each key point. Then for landmark normalization, the pose centres are moved to (0,0) and they are scaled to a constant pose size.

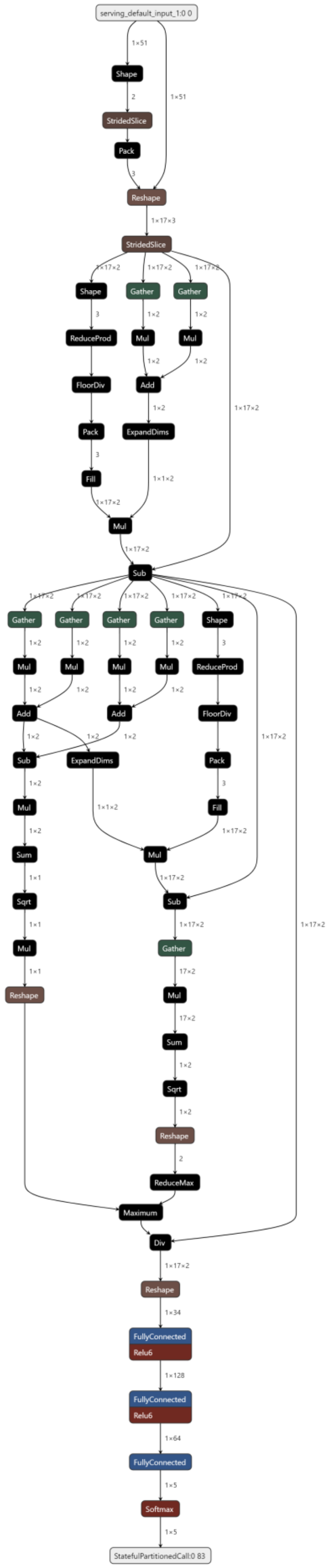

Fig. 3. BowlingDL model architecture

Finally, the normalized landmark coordinates are flattened into a vector and passed through multiple dense layers which produce an output of the shape 5 (Corresponding to the 5 class of bowling posture).

D. *BowlingDL Model Training and Results*

The BowlingDL was trained using the body key points data that was detected from the Bowling Posture dataset using the MoveNet model. The dataset was split into training (80%) and testing (20%) datasets. Table 3 lists the number of images for each category in the training and testing datasets. The training dataset was further split into training (85% of the training dataset) and validation (15% of the training dataset).

TABLE 3. TRAINING AND TESTING DATASETS

Class	Training	Testing
One-Handed Left	24	5
One-Handed Right	21	6
Two-Handed Left	18	5
Two-Handed Right	19	7
Improper	40	7
Total	**122**	**30**

Fig. 4 depicts the model accuracy during the training process for training and validation datasets. At the end of the 40th epoch, the model achieved 80% for training and 79% for validation data. The accuracy increases rapidly in the first 20 epochs, then it fluctuates between 70% and 80% in the last 20 epochs. This can be related to the size of our dataset. It is believed that by increasing the size of the dataset the model performance would improve.

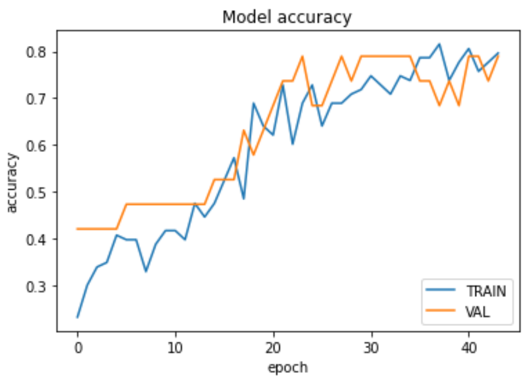

Fig. 4. The BowlingDL model accuracy

The BowlingDL model's performance was evaluated using the test dataset. The model achieved 83% accuracy on the test dataset. Fig. 5 illustrate the confusion matrix of the test dataset for the BowlingDL model. It compares the true labels of the test dataset with the BowlingDL model predicted labels for the five classes. The dark diagonal line shows the truly predicted classes by the BowlingDL model. Table 4 compares the accuracy values of all training, validation, and testing datasets.

TABLE 4. THE BOWLINGDL ACCURACY PERFORMACE

Dataset	BowlingDL Accuracy	
Training	0.7961	80%
Validation	0.7895	79%
Testing	0.8333	83%

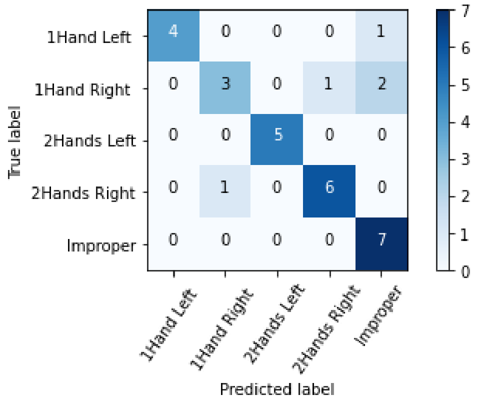

Fig. 5. Confusion matrix of the BowlingDL model

E. *BowlingDL Model Deployment*

For deployment and testing, an edge-friendly version of BowlingDL was generated using TensorFlow Lite (TFLite). The TFLite Converter library was used to convert the trained BowlingDL model into a TFLite model. In addition, a bowling smart mobile application was developed for bowling players using Android studio. The application continuously detects player poses in the frames fed by the device's camera and provides pose classification results. Algorithm 1 shows the procedure that the application follows to predict the type/class of a bowling posture. The MoveNet Thunder model was used to detect body parts (key points) and draw the pose skeleton on the frame. After that, the key points are fed into the BowlingDL TFLite model and the top three classification results are shown to the user along with their score. Fig. 6 shows a screenshot of the developed mobile application.

Algorithm 1: Bowling Pose Classification

```
Input: bowling_posture_image
Output: result[3][2]  // a sorted list of the top three classifications
Function: get_player_pose (bowling_posture_image)
    Init: size ← model input dimension,
          movenet ← Movenet ('movenet_thunder')
    //Image pre-processing
    posture_image ← posture_image.resize(size, size)
    //Detect player pose using MoveNet
    keypoints = movenet.detect(posture_image)
    posture_image← draw_skeleton (posture_image,keypoints)
    //Classification
    model ← load_model('BowlingDL') //load trained model
    P[p_0, ... , p_C] ← model.predict(keypoints)
      //Find the top three probability values
    For i in 3
        prediction_class ← 0
        probability ← p_0
        For c in C   //C is # bowling posture classes
            If  P[c] > probability Then
                probability ← P[c]
                prediction_class ← c
            End If
        End For
        result.add(prediction_class, probability)
    End For
    Return result
```

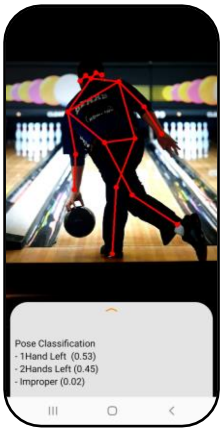

Fig. 6 BowlingDL mobile application

V. Conclusion

HPE has gained a lot of attention due to its versatile potential applications in various domains including transportation, healthcare, gaming, augmented reality, and sports. In sports, HPE can be used to build smart sports applications for learning, evaluation, and coaching. This paper proposed a deep-learning approach for bowling players' pose estimation and classification. The MoveNet model was used to detect various key points in bowling players' poses and our BowlingDL model was used to classify key points into five different classes of bowling players' poses. Our proposed model achieved 80% accuracy for the training dataset and 83% for the testing dataset. In addition, a smart mobile application for bowling players was developed where an edge-friendly version of BowlingDL was deployed.

One of the main challenges is creating a dataset for bowling posture as no image dataset was found for bowling posture. Therefore, in future, a larger dataset with variant classes of bowling player poses will be considered to improve the Bowling Posture Dataset and the BowlingDL model performance. In addition, other techniques such as semantic search and deep search might be adopted.

References

[1] T. L. Munea, Y. Z. Jembre, H. T. Weldegebriel, L. Chen, C. Huang, and C. Yang, "The Progress of Human Pose Estimation: A Survey and Taxonomy of Models Applied in 2D Human Pose Estimation," *IEEE Access*, vol. 8, pp. 133330–133348, 2020, doi: 10.1109/ACCESS.2020.3010248.

[2] M. Cormier, A. Clepe, A. Specker, and J. Beyerer, "Where are we with Human Pose Estimation in Real-World Surveillance?," *Proc. - 2022 IEEE/CVF Winter Conf. Appl. Comput. Vis. Work. WACVW 2022*, pp. 591–601, 2022, doi: 10.1109/WACVW54805.2022.00065.

[3] S. N. Boualia and N. E. Ben Amara, "Deep full-body HPE for activity recognition from RGB frames only," *Informatics*, vol. 8, no. 1, p. 2, Jan. 2021, doi: 10.3390/informatics8010002.

[4] A. Raj, D. Singh, and C. Prakash, "Active human pose estimation for assisted living," *ACM Int. Conf. Proceeding Ser.*, pp. 110–116, Aug. 2021, doi: 10.1145/3474124.3474139.

[5] Z. Fang and A. M. Lopez, "Intention Recognition of Pedestrians and Cyclists by 2D Pose Estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 11, pp. 4773–4783, Nov. 2020, doi: 10.1109/TITS.2019.2946642.

[6] A. Moryossef, I. Tsochantaridis, R. Aharoni, S. Ebling, and S. Narayanan, "Real-Time Sign Language Detection Using Human Pose Estimation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12536 LNCS, pp. 237–248, 2020, doi: 10.1007/978-3-030-66096-3_17/COVER.

[7] L. Kumarapu and P. Mukherjee, "AnimePose: Multi-person 3D pose estimation and animation," *Pattern Recognit. Lett.*, vol. 147, pp. 16–24, Feb. 2020, doi: 10.48550/arxiv.2002.02792.

[8] A. Nadeem, A. Jalal, and K. Kim, "Automatic human posture estimation for sport activity recognition with robust body parts detection and entropy markov model," *Multimed. Tools Appl. 2021 8014*, vol. 80, no. 14, pp. 21465–21498, Mar. 2021, doi: 10.1007/S11042-021-10687-5.

[9] Y. Xu and Y. Peng, "Real-Time Possessing Relationship Detection for Sports Analytics," *Chinese Control Conf. CCC*, vol. 2020-July, pp. 7373–7378, Jul. 2020, doi: 10.23919/CCC50068.2020.9189516.

[10] J. Wang, K. Qiu, H. Peng, J. Fu, and J. Zhu, "AI Coach: Deep Human Pose Estimation and Analysis for Personalized Athletic Training Assistance," Oct. 2019, pp. 374–382. doi: 10.1145/3343031.3350910.

[11] J. Zou *et al.*, "Intelligent fitness trainer system based on human pose estimation," *Lect. Notes Electr. Eng.*, vol. 550, pp. 593–599, 2019, doi: 10.1007/978-981-13-7123-3_69/COVER.

[12] A. Badiola-Bengoa and A. Mendez-Zorrilla, "A Systematic Review of the Application of Camera-Based Human Pose Estimation in the Field of Sport and Physical Exercise," *Sensors 2021, Vol. 21, Page 5996*, vol. 21, no. 18, p. 5996, Sep. 2021, doi: 10.3390/S21185996.

[13] A. Arbues-Sanguesa, A. Martin, J. Fernandez, C. Rodriguez, G. Haro, and C. Ballester, "Always Look on the Bright Side of the Field: Merging Pose and Contextual Data to Estimate Orientation of Soccer Players," *Proc. - Int. Conf. Image Process. ICIP*, vol. 2020-October, pp. 1506–1510, Oct. 2020, doi: 10.1109/ICIP40778.2020.9190639.

[14] M. Sypetkowski, G. Sarwas, and T. Trzciński, "Synthetic Image Translation for Football Players Pose Estimation," *JUCS - J. Univers. Comput. Sci. 25(6) 683-700*, vol. 25, no. 6, pp. 683–700, 2019, doi: 10.3217/JUCS-025-06-0683.

[15] T. Shimizu, R. Hachiuma, H. Saito, T. Yoshikawa, and C. Lee, "Prediction of future shot direction using pose and position of tennis player," *MMSports 2019 - Proc. 2nd Int. Work. Multimed. Content Anal. Sport. co-located with MM 2019*, pp. 59–66, Oct. 2019, doi: 10.1145/3347318.3355523.

[16] S. Garg, A. Saxena, and R. Gupta, "Yoga pose classification: a CNN and MediaPipe inspired deep learning approach for real-world application," *J. Ambient Intell. Humaniz. Comput. 2022*, pp. 1–12, Jun. 2022, doi: 10.1007/S12652-022-03910-0.

[17] S. Kothari, "Yoga Pose Classification Using Deep Learning," *Master's Proj.*, May 2020, doi: https://doi.org/10.31979/etd.rkgu-pc9k.

[18] G. M. Difini, M. G. Martins, and J. L. V. Barbosa, "Human Pose Estimation for Training Assistance: A Systematic Literature Review," *ACM Int. Conf. Proceeding Ser.*, pp. 189–196, Nov. 2021, doi: 10.1145/3470482.3479633.

[19] J. Wang *et al.*, "Deep 3D human pose estimation: A review," *Comput. Vis. Image Underst.*, vol. 210, p. 103225, Sep. 2021, doi: 10.1016/J.CVIU.2021.103225.

[20] "YouTube." https://www.youtube.com/ (accessed Oct. 07, 2022).

[21] "PBABowling - YouTube." https://www.youtube.com/c/pba (accessed Oct. 07, 2022).

[22] "National Bowling Academy - YouTube." https://www.youtube.com/c/NationalBowlingAcademy (accessed Oct. 07, 2022).

[23] "InsideBowling - YouTube."

https://www.youtube.com/c/insidebowling (accessed Oct. 07, 2022).
[24]“The House Bowling - YouTube.” https://www.youtube.com/c/TheHouseBowling (accessed Oct. 07, 2022).
[25]“MoveNet: Ultra fast and accurate pose detection model. | TensorFlow Hub.” https://www.tensorflow.org/hub/tutorials/movenet (accessed Oct. 07, 2022).
[26]“TensorFlow.” https://www.tensorflow.org/ (accessed Oct. 07, 2022).