Shifting the Weight: Applications of AI in Olympic Weightlifting

Daniel Bolarinwa
Engineering and Computing
University of East London
London, United Kingdom
dtireni@gmail.com

Dr Nadeem Qazi
Engineering and Computing
University of East London
London, United Kingdom
n.qazi@uel.ac.uk

Dr Mustansar Ghazanfar
Engineering and Computing
University of East London
London, United Kingdom
m.ghazanfar@uel.ac.uk

***Abstract*—The role of humans in time-pressured decision-making processes within sports has been critically examined in psychological research. This is particularly relevant in complex movement sports such as Dressage, Gymnastics, and Olympic Weightlifting. Not only are humans susceptible to bias, but they also lack the necessary processing capacity to assess intricate movements in real-time. Although some research has been conducted in this space very few use Computer Vision based approaches. To address this issue, this research proposes a novel Computer Vision solution to automate the judging process in Olympic Weightlifting. The solution incorporates LSTM-based Gesture Recognition and Human Pose Estimation using Mediapipe. The feasibility and effectiveness of the proposed solution are assessed by leveraging a combination of videos from the official Olympics YouTube channel and amateur recorded videos captured from the perspective of the Olympic Weightlifting Centre judge. The findings indicate a high degree of success in achieving the research objective. The solution achieved a validation accuracy of 96% and an average F1 score of 0.91. These results demonstrate the plausibility and efficacy of the proposed approach in automating the judging process within Olympic Weightlifting. By automating this process, the potential influence of human bias can be mitigated while improving the real-time assessment of complex movements. The implications of these findings extend beyond Olympic Weightlifting and have the potential to enhance judging processes in other complex movement sports as well.**

Keywords*—Olympic Weightlifting, Gesture Recognition, Human Pose Estimation, LSTM, Bias, Judge/s, Snatch, Clean and Jerk

I. Introduction

The exciting and dynamic sport of Weightlifting decides its winners through allegedly 'objective' judge interpretations and analysis of an athlete's movement. The International Weightlifting Federation (IWF) defines extremely detailed rules and regulations around what counts as correct and incorrect movement. Nonetheless, this judging system can be flawed because it is ultimately left to human interpretation of the rules. This poignant flaw introduces room for bias-induced decision-making. We can assume that judges always try to remain fair during a competition. Yet considerable evidence exists within sports psychological research that "getting it right" isn't as easy as it sounds [1].

In highly complex movement sports, the demands on judges' information processing capacities far outweigh what can reasonably be expected, considering time and social pressures [1]. As a result, judges tend to fall back on bias-induced schemas, including patriotism bias which suggests a judge favouring an athlete from their own country, reputation bias which suggests influence based on the athlete's reputation, rank order bias which refers to the tendency to expect a good or bad performance based on which stage of the competition the performance takes place, conformity effect bias which suggests that when judges can see the scores given by their judging peers, either during or at the end of a competition, they are likely to adapt their scoring to "fall in line". As a result, accurate performance evaluation within complex movement sports like Weightlifting is a challenge.

When examining prior attempts to address bias in the judging process of complex movement sports, there is a scarcity of bespoke technological solutions that directly tackle the problem. Even among the solutions that do exist, judges continue to play a central role, making them susceptible to bias-induced flaws in high-pressured environments similar to sport judging [1].

This research proposes a novel Computer Vision solution with two aims: (1) identify as objectively as possible if a weightlifting movement can be classified as successful; (2) provide the rationale behind the decision that was made.

II. Related Research

This section dives deeper into the psychology behind biased decision-making and its prevalence in elite-level sports as well as previous solutions which have attempted to tackle bias-induced action quality assessment in sports.

A. Human Psychology Behind Bias in Sports Judging

Judges enforce the rules of a sport, however in moments of controversy when critical split-second decisions need to be made, the "human" element of refereeing comes to the fore [1]. If judging mistakes occur, the effects are as significant as an athlete winning or ending a competition with nothing.

A true evaluation of what constitutes bias in human psychology is required to understand why it would occur in the first place. As we proceed to do so, questions arise – what are the variables within a human's thought process before making a subconsciously biased decision?

Perception is the first cognitive step when a person makes a judgement. Perception is heavily influenced by the prior knowledge an individual has accrued [1]. In Weightlifting, this would be a judge's perception of an athlete's movement. After gaining perception, a judge must assign meaning to the performance they have witnessed by drawing on previous memories – have they seen this movement before, have they seen this athlete perform before, does the athlete have a good track record of performing well, what types of performances have previously been considered good or bad? All these questions play major roles in the output decision.

When assessing performances, judges are required to provide their verdict under increasing time and social pressure to "get it right". This can lead to judges taking shortcuts to get to their decisions.

The snatch and clean and jerk contain several technical and artistic elements which all need to be evaluated at once. However, processing of such complex information exceeds human capabilities. To conform to the social pressures, timeframes, and expectations judges fall back on these shortcuts which help them come up with a judgement that in their mind accurately approximates actual performance levels.

B. *Forms of Bias in Elite-Level Sports*

Similar to fans, judges may overlook flaws in the performance of a well-known athlete, such as a serial world champion or Olympic gold medalist. For example, at the London 2012 Olympics, Japanese gymnast Kohei Uchimura was widely considered the greatest male artistic gymnast of all time [2], earning himself the nickname 'the king' among Japanese nationals [3]. During the men's team event, Uchimura made an uncharacteristic error during his dismount on the pommel horse apparatus, resulting in a lower score of 13.466 and a fourth-place finish for the Japanese team. However, the Japanese coaching staff filed a request for a re-evaluation, which was accepted. 'King' Kohei's score was subsequently increased to 14.166, moving the Japanese team up to second place surpassing Ukraine and Great Britain [4]. It can be deduced that Uchimura's reputation played a significant role in the decision to re-evaluate his performance. This begs the question – would the same have happened if it was a lesser-known athlete?

Another example is conformity bias, which is especially poignant in a system where a judge can see other judges' verdicts. This can be key in Weightlifting as all judges can see the light which indicates the verdict of the others. As a result, a judge is likely to modify their scoring for it to fall in line with the decisions of the majority. If a judge missed a part of a performance due to a lapse in concentration, they would draw on the decisions made by other judges to come to a decision of their own.

C. *Existing Solutions to Tackle Bias in Sport Judging*

1) *Using Statistics to Judge the Judges*

Heiniger and Mercier [5] implemented a statistical engine to analyse the performance of gymnastics judges with three objectives: (1) provide constructive feedback to judges, executive committees, and national federations; (2) assign the best judges to the most important competitions; and (3) detect bias and persistent misjudging related to patriotism bias.

A model was developed to generate a judge's marking score which scales the difference between the mark of a judge and the true performance level of a gymnast as a function of the intrinsic judging error variability estimated from historical data for each apparatus through Equation (1). Where $e_{p,j}$ is the true judging error of judge j for performance p, $\hat{\sigma}_d(c_p)$ is the intrinsic judging error variability of discipline (apparatus) d, $\delta_{p,j}$ is the judging discrepancy of judge j for performance p, $s_{p,j}$ is the actual mark given by judge j for performance p, and c_p is the control score obtained for performance p. The marking score $m_{p,j}$ quantifies the accuracy of a judge compared to their peers. The marking score of a perfect judge in this case would be 0.

$$m_{p,j} \triangleq \frac{e_{p,j}}{\hat{\sigma}_d(c_p)} \approx \frac{\delta_{p,j}}{\hat{\sigma}_d(c_p)} = \frac{s_{p,j} - c_p}{\hat{\sigma}_d(c_p)} \qquad (1)$$

2) *Hawk-Eye*

Hawk-Eye is a vision system which traces an object's trajectory during any given period. It is currently one of the most advanced officiating tools used across many sports like Tennis, Football, Cricket, Volleyball, Ice Hockey, Horse Racing and even NASCAR [6]. In Tennis, Hawk-Eye's ultra-motion cameras can work up to 340 frames per second to render the trajectory, bounce mark, and contact areas of a tennis ball to real video footage for instant feedback use when an athlete challenges a judge's line call [6].

It is important to know that applications of such technologies also provide athletes with a clear explanation for given decisions therefore reducing disruptions in competitions due to player uproar and challenges. Former female world No. 1 Maria Sharapova stated "As a player, I want to know the line calls are as accurate as technology will allow. In that sense, [Hawk Eye] is great news for all players." [7], further proving benefits to using technology in reducing errors in judge decision making.

D. *Computer Vision Applications in Weightlifting*

A key area of Computer Vision that can be utilised to tackle the problem statement is Human Pose Estimation. Human Pose Estimation refers to the process of detecting the location of a person within an image by recognizing, locating, and tracking key points on a person [8]. As a result, Human Pose Estimation can have 2 benefits within weightlifting: (1) it can be applied for human action quality assessment as we can train neural networks from identified poses to the IWF's rules used by judges for athlete performance evaluation; (2) Human Pose Estimation can also provide a detailed feedback system to athletes on how they can improve parts of their movements and address any inefficiencies to polish up their technique for improved future performance.

Human action quality assessment is activity classification at its core and activity classification is itself a time series problem. Time-series classification is a type of supervised machine learning that is used to predict future values from past data using statistical techniques [9]. Neural networks have proven to be the most effective in achieving this.

1) *Human Activity Recognition from sensor data using Deep Neural Network*

Clouthier et al. [10] aimed to use deep learning techniques to automatically identify movements typically found in movement screens and assess the feasibility of performing the classification based on wearable sensor data. Movement screens are used to assess the overall movement quality of an athlete.

Data-driven approaches have the potential to improve the repeatability of scoring and increase the ability to detect subtle differences in movement patterns [10]. The use of wearable sensors was an attractive alternative to optical motion capture for motion analysis applications. Reference [10] had 2 aims; (1) to use a deep neural network to identify when movements typical of movement screens occur within motion data; (2) to compare networks trained using optical motion capture data with those trained using data available from wearable sensors.

The idea that the combination of CNNs to extract features with long short-term memory (LSTM) recurrent networks to capture temporal dependencies would improve classification performance over CNNs alone was explored.

The results of this research achieved high classification accuracy scores with an F1 score of ≈0.90.

2) Objectively Measuring Athlete Performance in Olympic Weightlifting

Karunaratne et al. introduced a method to assess athlete techniques in weightlifting by using skeleton-based human action recognition [11].

Concepts from [12] formed a foundation in identifying the crucial information required to assess the quality of weightlifting movements. OpenPose was used to extract the athlete key points within the movement as opposed to a Kinect sensor. This means any camera can be used to capture the athlete's movement enabling them to implement the same equations proposed by [12] to calculate barbell velocity, barbell angles, and knee joint angles on the OpenPose key points.

A scoring model was proposed using the Multilayer Perceptron feed-forward neural network from Scikit-learn built on top of the OpenPose CMU model - a 2D pose estimation model.

This proposed model achieved a test data classification accuracy of 93% using publicly available data from the official Olympics YouTube channel of the 2016 Olympics.

E. Summary and Gap Analysis

After dissecting previous research on bias within judging systems in sports and the proposed solutions to mitigate it, it becomes evident that a universal solution is not achievable.

In Weightlifting, very little previous technologies have plausible implementations. For example, the research conducted in [10] requires data being extracted from sensors that are attached to athletes for activity recognition. However, the need to attach sensors to athletes may have negative effects on an athlete's performance. They may feel uncomfortable and obstructed by the sensors. It would also introduce an extra process of setting up the sensors leading up to an athlete making their attempts which can also affect their mental state before conducting their lift.

Chatzitofis et al. [12] also suggested that a weightlifting attempt is considered successful when the bar is on top of the athlete, and she/he keeps it balanced for some time. This statement is flawed because there are many points of failure during a lift despite reaching the end position of having the barbell overhead in a stable position which would most definitely result in the movement being classified as unsuccessful. An example of this is during the catch position of a clean, if the athlete's arms contact the legs despite getting up with the barbell overhead the lift is classed as unsuccessful because the IWF defines '*Touching the thighs or the knees with the elbows or the upper arms*' as prohibited [13].

This research aims to address the gaps in previous research by introducing a novel Computer Vision solution using a combination of 3D Human Pose Estimation and Gesture Recognition techniques. This method will aim to mitigate the effects of bias within the weightlifting judging process by classifying whether an athlete's attempt at completing the weightlifting movements is successful or unsuccessful.

III. Implementation

All implementation steps were conducted on the following hardware architecture: an AMD Ryzen 5600X CPU overclocked from base 3.9GHz to 4.4 GHz, 16GB of GDDR6 RAM, and NVIDIA GeForce RTX 3060 Ti GPU.

A. Dataset

We conducted an extensive search for publicly available weightlifting video/image datasets but none existed. We created our datasets using a combination of videos from official Olympic Weightlifting competitions (ranging from London 2012 to Tokyo 2022) and amateur-recorded videos of consenting gym-goers. The final video sample size was 858 videos which amounted to ≈180,000 frames. To ensure consistency across the dataset, all video recordings were captured with a focus on the centre judge's viewpoint and a resolution of 1080p with a frame rate of 30 fps.

We determined that there were 4 keys classes for a Weightlifting movement: (a) complete snatch, (b) complete clean and jerk, (c) incomplete snatch, (d) incomplete clean and jerk. Where 'complete' is defined by the IWF as 'the lifted weight must be maintained in the final motionless position, with the arms and legs fully extended and feet on the same line and parallel to the plane of the trunk and the barbell' [13].

Fig. 1 illustrates the 4 aforementioned classes where *complete snatch* is identified by the recovery phase poses that are highlighted by the red box. The same applies for the *complete clean and jerk.* If these highlighted poses are not identified in a video sequence, then the video is classified as *incomplete snatch* or *incomplete clean and jerk* respectively.

Fig. 1. Illustration depicting the 4 classes mentioned above. Where the complete movement classes are hightlighted in red boxes relative to the given movement.

One of the key raw data preprocessing steps included rescaling the videos' lengths to a frame sequence length of 75 for efficient feature extraction and improved model training times. This was achieved using a simple algorithm derived from a custom code script described in Equation (2), where the original video of length n is denoted as x and the desired output length as y. If n > y, we skip frames in intervals determined by the floor division of n by y, otherwise, the original video is returned.

$$f(x_n, y) = \begin{cases} x_i \text{for i in } \left[0, \left\lfloor\frac{n}{y}\right\rfloor, \dots, n\left(\left\lfloor\frac{n}{y}\right\rfloor\right)\right], & n > y \\ x, & \text{otherwise} \end{cases} \quad (2)$$

After raw video data preparation and pre-processing, we applied transfer learning using the TensorFlow InceptionV3 CNN [14] to extract relevant features from each video. We processed each frame from the newly rescaled videos while preserving their sequential time-series nature. The output from the final pooling layer of the network produced a 2048-dimensional vector of features which is then compiled for each video into a final sequence to form the input for the subsequent classification task.

Fig. 2 and 3 show the dataset distribution for our 4 classes in both training and validation dataset splits.

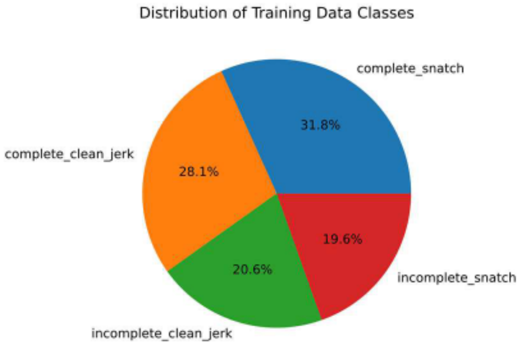

Fig. 2. Training Data Class Distribution – highlighting the spread of the training data between the identified Olympic Weightlifting movement classes

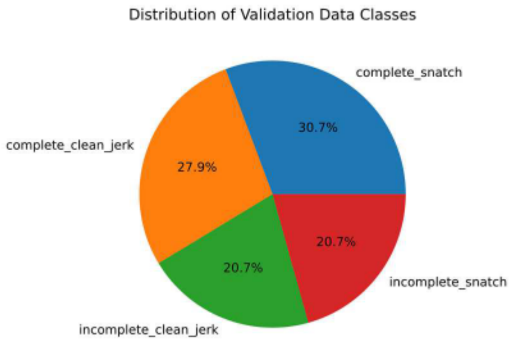

Fig. 3. Validation Data Class Distribution - highlighting the spread of the validation data between the identified Olympic Weightlifting movement classes

The final dataset split was 80/20 for training and validation respectively with a further 80/20 split of the training data to form a train, test, validation split of 64%, 16%, 20% respectively.

B. Gesture Recognition with a Long Short-Term Memory Recurrent Neural Network

We implemented a shallow RNN with a single LSTM layer and a dense layer with dropouts in between. LSTM was the preferred choice due to its ability to capture temporal dependencies in sequential data. Furthermore, LSTM has a memory cell that can store and propagate information over time, this enables the network to remember important context from earlier frames even when there is a time gap between relevant actions. This helps to capture long-term dependencies between video frames, which are crucial for accurate recognition of complex activities.

To improve the model results, we used KerasTuner - a scalable hyperparameter optimization framework that simplifies the usually cumbersome task of hyperparameter tuning [15]. We focused on KerasTuner's GridSearch method, where we provided search spaces and thresholds for key parameters in our network, such as the LSTM units, dense units, dropout rate, learning rate, decay, and regularisation rate. KerasTuner then trained models on every possible combination of hyperparameters within the provided search space, using a subset of our data.

The best model based on model accuracy and loss, along with the hyperparameters used for that model was – 1024 LSTM units with an initial drop-out rate of 0.2, 128 dense units, an L2 regulariser with a penalty coefficient of 0.001, a second drop-out layer with a rate of 0.2, and an optimizer with a learning rate of 0.00001 and decay of 0.0001. We retrained the final model on our entire dataset to ensure its robustness.

The resulting network architecture depicted in Fig. 4, showcases an input dimension of (75, 2048) to represent the 2048 feature vectors for every rescaled 75-frame video. It also shows the dense layer output which employs the softmax activation function, enabling us to derive a probability value for the video's classification into each of the 4 gesture classes.

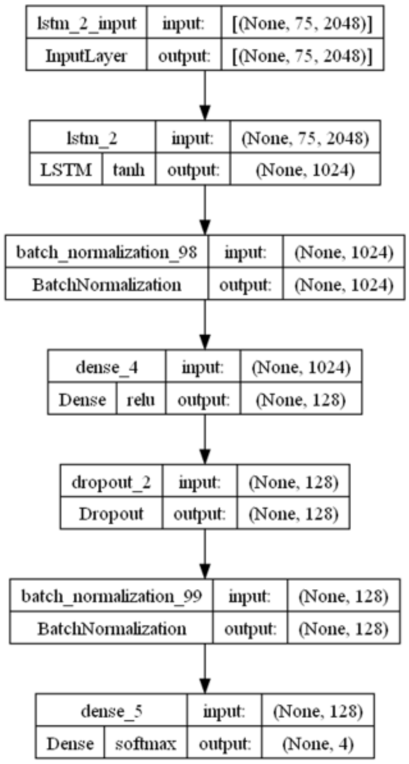

Fig. 4. LSTM Architecture – defining structure and parameters of the LSTM network including dimensionality changes after each layer

The LSTM model used in our study incorporates several noteworthy features, some of which are not captured in Fig. 3. One of these features is L2 regularization, also known as Ridge Regression. This technique is used to prevent overfitting by adding penalties to the loss function of the model. The penalty is proportional to the sum of the squares of the weights in the model, as shown in Equation (3) where MSE (Mean Squared Error) is a measure of the difference between the model's predicted and actual values, ypred is the predicted values from the model, yorig is the actual values from the dataset, n is number of samples in the dataset, λ is the regularisation rate, and mi is the weight parameter in the model.. L2 regularisation encourages the model to learn simpler weights that are closer to zero allowing the model to better generalise to unseen data.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_{\mathrm{pred}} - \mathrm{Y}_{\mathrm{orig}}\right)^2 + \lambda\sum_{i=1}^{n} \mathrm{m}_i^2 \qquad (3)$$

C. Human Pose Estimation with OpenCV and Mediapipe

Mediapipe uses a pose graph algorithm to connect body parts based on body part spatial relationships and human anatomy constraints which can then be overlayed on an input image. The resulting coordinates are 4 dimensional including the x for horizontal representation, y for vertical representation, z for depth representation, and v to represent the visibility of the body part in the image [16]. The landmarks generated by Mediapipe can be found in Fig. 5. Pose Estimation is paramount for this research because it allows the system to accurately track and analyse the precise positioning of an athlete's body during weightlifting movements. This level of granularity provides an objective assessment of form, balance, and overall execution which gesture recognition alone cannot achieve.

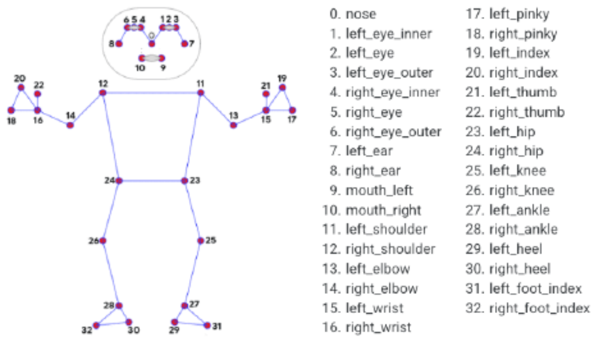

Fig. 5. Mediapipe Body Landmarks – shows all the landmarks that are available to conduct biomechanical calculations with and the index to use in order to access their coordinates

For the IWF rules we assessed, the important variables we needed to keep track of were: (1) the velocity of the athlete's hands/wrists throughout the lifts; (2) their elbow, shoulder, knee joint angles; (3) depth difference of the feet; (4) knee and elbow joint proximity; (5) relative travelled distance of the hip, wrist and shoulder; and (6) the time difference of which they held specific poses. The formulas used to calculate the joint angles and velocities are given by Equations (4) and (5) respectively.

$$\theta = \arccos\left(\frac{\mathbf{AB}\cdot\mathbf{BC}}{\|\mathbf{AB}\|\|\mathbf{BC}\|}\right) \tag{4}$$

In Equation (4), given 3 cartesian coordinates A, B, C, the joint angle θ of $A \rightarrow B \rightarrow C$ is the inverse cosine function of the dot product of the vectors $\overrightarrow{AB}$, $\overrightarrow{BC}$ divided by the product of the vector's magnitudes.

$$\text{velocity} = \frac{\mathbf{landmark_{yn}} - \mathbf{landmark_{yn-10}}}{\mathbf{time}} \tag{5}$$

Equation (5) is used to calculate landmark velocity given the distance travelled by the landmark in between 10 frame intervals. Given that the Olympic lifts occur around the body's sagittal plane we only considered the displacement of the y coordinates of the landmarks.

The final experimental design flow is seen in Fig. 6.

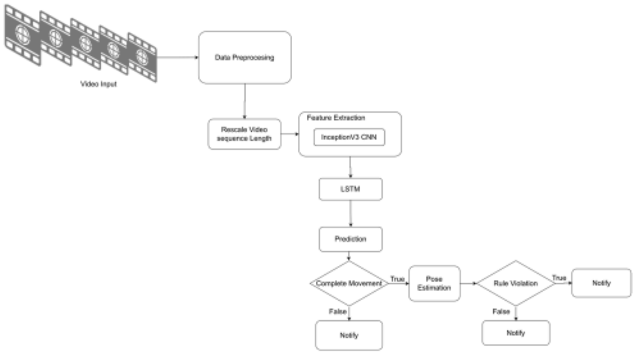

Fig. 6. Final Experimental design flow – high level design flow detailing the solution overview from data input to final pose estimation process

IV. Results

We will focus on the performance of the solution in identifying the 4 Weightlifting movement classes.

Fig. 7 shows the model's learning curve after training for 150 epochs (a complete iteration of the dataset). The model achieved an accuracy of 96% on the test subset of the data whilst achieving a loss value of 0.39. The dark blue line represents training data, the light blue line represents test data.

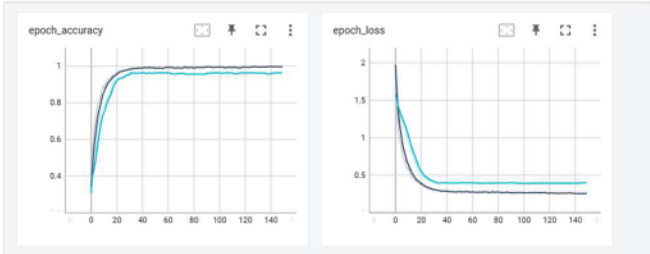

Fig. 7. LSTM Model Learning Curve Accuracy vs Loss – x-axis is the number of epochs the model was training for and the respective y-axes are the model's accuracy and loss values.

Fig. 8 shows the classification report when comparing the expected validation dataset labels to the predictions made by the model. The classification report provides a summary that includes the model's prediction precision, recall, and f1 scores. Precision scores measure how well the model avoids false positive predictions. Recall scores measure how well the model can make true positive predictions. F1 score is the balance between the precision and recall scores.

	precision	recall	f1-score	support
complete_clean_jerk	0.96	0.98	0.97	50
complete_snatch	0.98	1.00	0.99	55
incomplete_clean_jerk	0.97	0.78	0.87	37
incomplete_snatch	0.86	0.97	0.91	37
micro avg	0.94	0.94	0.94	179
macro avg	0.94	0.93	0.93	179
weighted avg	0.95	0.94	0.94	179
samples avg	0.94	0.94	0.94	179

Fig. 8. Classification Report – providing granular information on the LSTM's classification performance with respect to each class of the data

Fig. 9 shows the Receiver Operating Characteristic (ROC) Area Under the Curve (AUC) graph. ROC AUC graphs are largely a graphical representation of the classification report's information. It notably focuses on the correlation between the model's true positive rate and false positive rate for each class. This graph ties in well with Fig. 10 which shows more granular information of the predictions the model made against the different classes through a confusion matrix.

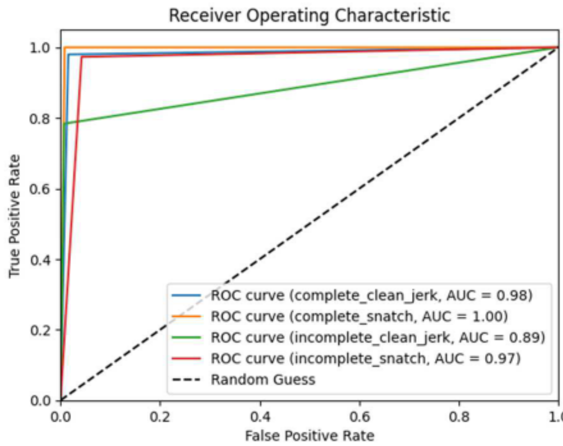

Fig. 9. ROC AUC Graph – calculated from the validation split of the data achieving by plotting the model's TPR against its FPR lowest score of 0.89 for the *incomplete_clean_jerk* class

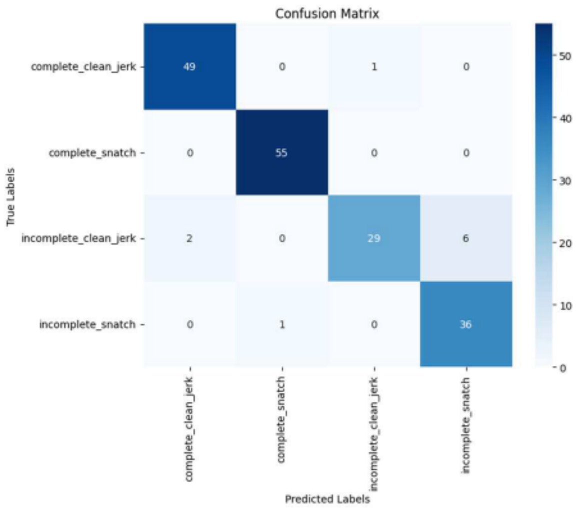

Fig. 10. Confusion Matrix marking the LSTMs performance comparing the data's true labels vs the model's predictions

V. Analysis and Evaluation

Despite the limited availability of data for the experiment, the LSTM-based Gesture Recognition model exhibited training, testing, and validation accuracies surpassing 90%. These results exceeded performance expectations considering data and compute constraints. Moreover, these results align with the results in Karunaratne et al. (2021) further emphasizing the effectiveness of employing LSTM for Gesture Recognition.

In this research, the official Olympic Weightlifting YouTube videos were employed as the ground truth, serving as a fundamental benchmark against which our model's performance was evaluated. The utilization of the official Olympic data as ground truth allowed us to establish a reliable and objective reference for the assignment of expected predicted gestures.

By adopting the official Olympic videos as the ground truth, we ensured the credibility and accuracy of our computer vision model's results. This approach enabled us to compare the model's outputs with the scores and assessments provided by expert human judges during the Olympic Weightlifting events. Consequently, we could assess the extent to which our automated judging system improved the human judging process.

It is worth noting that the reliance on official Olympic videos as the ground truth was complemented by the inclusion of amateur-recorded videos in our dataset. By incorporating videos from both official and amateur sources, we aimed to enhance the robustness and generalizability of our model, ensuring that it could effectively handle a diverse range of scenarios and settings commonly encountered in real-world judging situations.

Fig. 7 offers evidence of the model's near-optimal fit, supported by two key characteristics. Firstly, the training and validation loss curves consistently decrease and stabilize, indicating that the model has reached its training capacity. Additional iterations of the data would not substantially enhance accuracy or loss values. Secondly, the generalization gap, representing the disparity between training loss and validation loss, measured 0.15. The generalization gap serves as a critical indicator of the model's performance in predicting unseen data, with an optimal fit model ideally exhibiting a gap close to zero.

Although the model achieves high accuracy, a comprehensive analysis of Fig. 8, 9, and 10 uncovers certain limitations in the current solution. Notably, the recall score for the incomplete clean and jerk class is significantly lower compared to other classes in Fig. 8. This discrepancy arises from the inherent challenge of distinguishing between incomplete clean and jerk and incomplete snatch due to their similarities. To alleviate this classification confusion, extending the video sequence length beyond 75 frames would provide more contextual information and facilitate differentiation between the classes. Furthermore, Fig. 10 illustrates that the model erroneously identified 6 incomplete snatches as incomplete clean and jerks. To enhance the overall model's performance, additional data is required.

When assessing the Human Pose Estimation process, the results positively support the plausibility of the original objective of this research in automating action quality assessment. Fig. 11, 12, 13, 14, and 15 demonstrate an analysis of selected results for key movement rules outlined by the IWF.

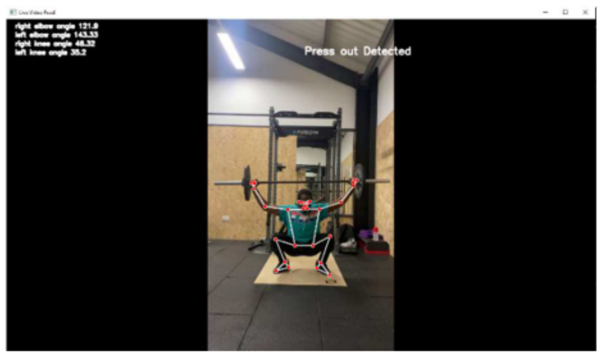

Fig. 11. Press out rule detection due to athlete elbow below ≈180°

Fig. 11 shows a snapshot of an instance where a press out was successfully detected by the system due to elbow joint angles of the athlete being below ≈180°. IWF defines 'Finishing with a press-out, defined as: continuing the extension of the arms after the athlete has reached the lowest point of his / her position in the squat or split for both the Snatch and the Jerk' [13] as incorrect movement.

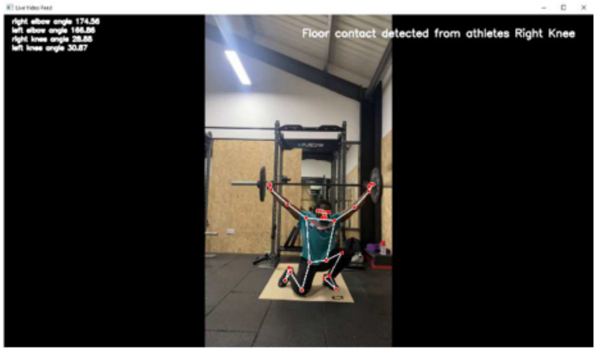

Fig. 12. Body to floor contact detection due to athlete knee contact with lifting platform at the bottom position of the snatch

Fig. 12 shows successful detection of body to floor contact as a result of the athlete's right knee touching the lifting platform. IWF defines 'Touching the platform with any part of the body other than the feet during the execution of the lift' as incorrect movement [13]. This positively demonstrates the powerful potential of using Mediapipe for action quality assessment.

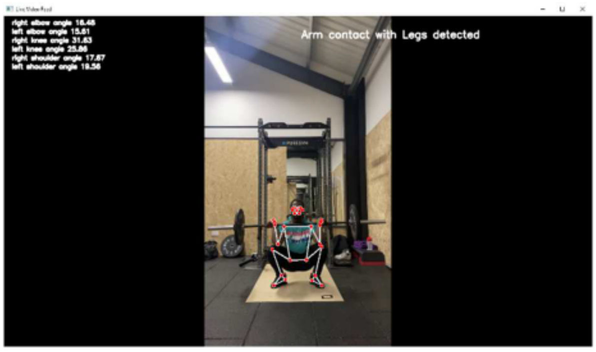

Fig. 13. Arms and Legs contact detection due to arm and leg contact at the bottom of the clean in a clean and jerk

Fig. 13 shows detection of contact between the arms and legs during the clean of an athlete's clean and jerk. IWF defines 'Touching the thighs or the knees with the elbows or the upper arms' as incorrect movement. This further demonstrates how effective the solution is at assessing different IWF movement rules.

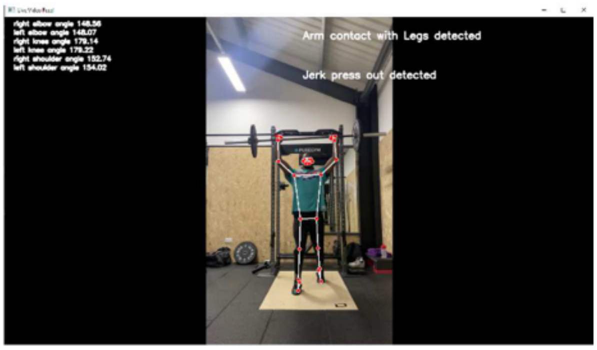

Fig. 14. Multiple incorrect movements detection simultaneously demonstrating a critical need to asess multiple rules at once

Fig. 14 depicts the ability of the solution to analyse and detect multiple incorrect movements simultaneously without any limitations on memory or storage capacity. In this case arm and leg contact was detected in the same movement prior to the snapshot detection of a jerk press out resulting in both errors being logged on screen.

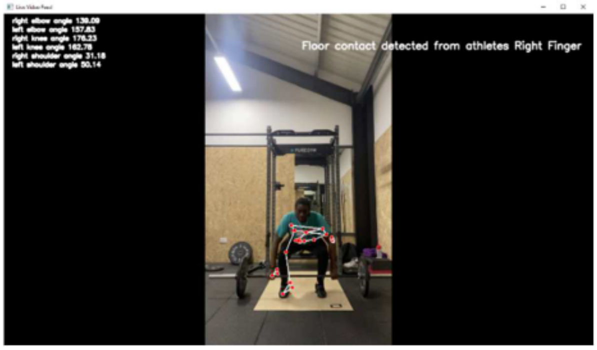

Fig. 15. False incorrect movement detection due to issues of clothing occlusion, and environment lighting

Fig. 15 shows an instance where the solution struggles to accurately identify all pose landmarks due to Mediapipe's inaccuracy when detecting landmarks in unconventional or challenging scenarios due to potential occlusion, clothing, or varying brightness settings. Misplacement or absence of certain pose landmarks can result in incorrect detection of rule violations as can be seen with the above example where the system incorrectly identified a violation of the 'floor contact' rule based on the athlete's right finger, despite the athlete not having initiated the lift yet.

To enhance this process, we can utilize higher frame rate videos recorded with higher resolution cameras similar to the high-speed cameras used within Hawk-Eye. We can also minimize motion blur and establish a standardised lifting environment where occlusion can be controlled, unlike in a commercial gym.

Despite some landmark mapping and estimation errors caused natively by the Mediapipe tool due to varying factors, Fig. 11, 12, 13, 14, and 15. showcase the potential and feasibility of our research in developing a robust and effective model for reducing bias-induced decision-making in the sports judging process, particularly in Olympic Weightlifting.

VI. Future Work and Conclusion

Given the psychological context behind what forms the basis of bias in decision-making, it is clear that relying solely on human judgment and evaluation is inadequate for guaranteeing fairness and precision in sports judging. This is not to say technology should replace human judges entirely, rather, technology can have a significant impact in expediting the elimination of bias within sports judging.

To enhance the proposed system, future endeavours should concentrate on expanding upon the neural network architectures already examined in this research. Specifically, addressing the notable limitations within the LSTM-based Gesture Recognition model in accurately predicting incomplete snatch and incomplete clean and jerk classes is crucial. This improvement can be achieved by dedicating efforts to collecting more data. This data may also incorporate the perspectives of the 2 side judges as our data focused only on the centre judge's view. Considering that the primary application of this solution would be within competitions, enriching the data in this manner is crucial.

Additionally, further enhancements to this research could involve evaluating the utilization of alternative 3D body pose estimation frameworks. This comparative analysis would enable a comprehensive assessment of which framework best aligns with the research objectives in accurately and objectively evaluating action quality even under challenging conditions such as poor lighting or occlusion.

Our research presented a Computer Vision solution that integrated Gesture Recognition and Human Pose Estimation, aiming to automate the judging process in Olympic Weightlifting. Though the proposed solution had limitations that prevented its readiness for production, it demonstrated significant success which signifies potential opportunities in this pursuit to pave the way for a fairer and more unbiased sports judging system.

REFERENCES

[1] Eurodressage (2010). Natural Bias, the Hidden Controversy in Judging Sports. [online] Eurodressage. Available at: https://www.eurodressage.com/2010/11/04/natural-bias-hidden-controversy-judging-sports

[2] UDDIN MUGHAL, K. (2015). Top 10 Male Gymnasts of All Time. [online] Sports Look. Available at: https://www.sportslook.net/top-10-male-gymnasts-of-all-time/ [Accessed 15 Jan. 2023].

[3] Soble, J. (2016). Japan Exults as Kohei Uchimura, Gymnast Called 'the King,' Expands His Empire. The New York Times. [online] 11 Aug. Available at: https://www.nytimes.com/2016/08/12/world/asia/japan-kohei-uchimura-olympics.html

[4] Zaccardi, N. (2012). Nick Zaccardi: Scoring controversy takes center stage at men's gymnastics final. [online] Sports Illustrated. Available at: https://www.si.com/more-sports/2012/07/30/2012-olympics-mens-gymnastics-final [Accessed 15 Jan. 2023].

[5] Heiniger, S. and Mercier, H. (2021). Judging the judges: evaluating the accuracy and national bias of international gymnastics judges. Journal of Quantitative Analysis in Sports, 17(4), pp.289–305. doi:https://doi.org/10.1515/jqas-2019-0113.

[6] Harrod Sport (2018). Hawk-Eye - Technology in Sport | Harrod Sport. [online] Harrodsport.com. Available at: https://www.harrodsport.com/advice-and-guides/hawkeye-technology-in-sport.

[7] Bal, B. and Dureja, G. (2012). Hawk Eye: A Logical Innovative Technology Use in Sports for Effective Decision Making. Sport Science Review, 21(1-2), pp.107–119. doi:https://doi.org/10.2478/v10237-012-0006-6.

[8] Zharovskikh, A. (2020). Pose Estimation - InData Labs AI Company Blog. [online] InData Labs. Available at: https://indatalabs.com/blog/pose-estimation [Accessed 5 Feb. 2023].

[9] Isakava, T. (2022). Human Activity Recognition: Everything You Should Know. [online] InData Labs. Available at: https://indatalabs.com/blog/human-activity-recognition.

[10] Clouthier, A.L., Ross, G.B. and Graham, R.B. (2020). Sensor Data Required for Automatic Recognition of Athletic Tasks Using Deep Neural Networks. Frontiers in Bioengineering and Biotechnology, 7. doi:https://doi.org/10.3389/fbioe.2019.00473.

[11] Karunaratne, A., Jayasooriya, C., Deegalla, S. and Navarathna, R. (2021). Objectively Measure Player Performance on Olympic Weightlifting. 2021 10th International Conference on Information and Automation for Sustainability (ICIAfS). doi:https://doi.org/10.1109/iciafs52090.2021.9605963.

[12] Chatzitofis, A., Vretos, N., Zarpalas, D. and Daras, P. (2013). Three-dimensional monitoring of weightlifting for computer assisted training. Proceedings of the virtual reality international conference: Laval virtual.

[13] IWF (2022). › IWF TECHNICAL AND COMPETITION RULES & REGULATIONS. [online] Available at: https://iwf.sport/wp-content/uploads/downloads/2022/11/IWF-TCRR-2022.pdf [Accessed 15 Jan. 2023].

[14] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2015). Rethinking the Inception Architecture for Computer Vision. doi:https://doi.org/10.48550.

[15] O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H. and Invernizzi, L. (2022). KerasTuner. [online] GitHub. Available at: https://github.com/keras-team/keras-tuner.

[16] Mediapipe. (2021). google/mediapipe. [online] Available at: https://github.com/google/mediapipe.

APPENDIX

Link to a demo of the final proposed solution can be found at https://clipchamp.com/watch/QGDo2XyvZTB