Can machine learning help reveal the competitive advantage of elite beach volleyball players?

Ola Thorsen[1], Emmanuel Esema[2], Said Hemaz[3],
Kai Olav Ellefsen[4], Henrik Herrebrøden[5], Hugh Alexander von Arnim[6], and Jim Torresen[7]

Abstract— **As the world of competitive sports increasingly embraces data-driven techniques, our research explores the potential of machine learning in distinguishing elite from semi-elite beach volleyball players. This study is motivated by the need to understand the subtle yet crucial differences in player movements that contribute to high-level performance in beach volleyball. Utilizing advanced machine learning techniques, we analyzed specific movement patterns of the motion of the torso during spikes, captured through vest-mounted accelerometers. Our approach offers novel insights into the nuanced dynamics of elite play, revealing that certain movement patterns are distinctly characteristic of higher skill levels. One of our key contributions is the ability to classify spiking movements at different skill levels with an accuracy rate as high as 87 %. This current research provides a foundation of what separates elite players from their semi-elite counterparts.**

I. Introduction

With the fast-paced evolution of sports analytics, using Artificial intelligence (AI) to accurately predict rally outcomes in beach volleyball can aid in developing strategies and enhance performance. AI-assisted analytics has the potential to reduce the workload of analysts and provide real-time, actionable insights for coaches and players. In beach volleyball, as in other sports, athletes display a rich diversity in physique, technique, and performance, shaping their unique approaches to the game.

Such variances present challenges when comparing and contrasting performances across athletes. The fluidity and versatility seen in beach volleyball mean that a single player might approach the same task differently at different times. One serve might be powerful and direct, while another could be deceptive and well-paced. This complexity, when coupled with the inherent differences between individual players, makes it difficult to draw straightforward correlations between technique and success. Variability underpins individual differences in motor control strategies and is vital for optimizing training methods [1]

However, developments in machine learning (ML) provide us with tools to gain insight into these nuances. Since early applications in sports during the late 20th century, ML-based analysis techniques have grown in sophistication and potential applicability. Yet, their penetration into certain sports such as beach volleyball remains nascent. Our work on the other hand represents the opposite [2]. Like less broadly popular sports such as table tennis and water polo, beach volleyball lacks significant attention from the betting market. As this market contributes to the drive towards advances in performance prediction [3], there is a significant gap in the literature specifically relating to the application of ML techniques in the analysis of beach volleyball performance. This paper aims to bridge this gap, exploring the potential of ML in distinguishing between elite and semi-elite beach volleyball players based on movement data.

Our central goals are: 1) To discern key differences in movement patterns between elite and semi-elite beach volleyball players using ML techniques applied to data from torso-mounted accelerometers. 2) To investigate which aspects of the game most distinctly delineate elite from semi-elite players, we chose side-out (which encompasses pass, set, and spike). The side-out phase is when the ball is received (passed) after the opponent's serve, after which the ball is set and spiked over the net 3) To explain and interpret the results we get from the trained model when it comes to what motion, or what part of the game, most clearly separates the two groups of players.

Fig. 1: Beach volleyball training

II. Related work

A. General Trends in Sports Analysis

Data-driven approaches and ML in sports analysis have been gaining momentum. A comprehensive literature review

*This work is supported by the Research Council of Norway through its Centres of Excellence scheme, project number: 262762

[1]O. Thorsen is with the Department of Informatics, University of Oslo, Oslo, Norway olathor@ifi.uio.no

[2]E. Esema is with the Department of Informatics, University of Oslo, Oslo, Norway emmanuje@ifi.uio.no

[3]S. Hemaz is with the Department of Informatics, University of Oslo, Oslo, Norway saidhem@ifi.uio.no

[4]K. O. Ellefsen is with the Department of Informatics, University of Oslo, Oslo, Norway kaiolae@ifi.uio.no

[5]H. Herrebrøden is with the School of Health Sciences, Kristiania University College, Oslo, Norway henrik.herrebroden@kristiania.no

[6]H. A. von Arnim is with RITMO, University of Oslo, Oslo, Norway hughav@imv.uio.no

[7]J. Tørresen is with the Department of Informatics and RITMO, University of Oslo, Oslo, Norway jimtoer@ifi.uio.no

by Rajšp [3] highlights the growing use of advanced data analysis techniques in sports. This systematic review of 109 studies emphasizes the significance of harnessing advanced data analysis techniques: Support Vector Machines (SVM) and Neural Networks were used in 19 and 14 studies, respectively. Focus on Popular Sports, such as soccer (12 studies) and running (11 studies), emphasizes the application of AI in more globally recognized sports.

AI and data mining are increasingly used for extracting practical knowledge from vast amounts of data, with sports like cross-country skiing [4], roller ski skating [5], and overground running [6] following this trend. However, due to the focus on individual sports in sports sciences, beach volleyball has seen limited advances in and also because of the relatively recent adoption of sensor technology in connection with the world championship in Vienna in 2017. There is therefore much potential for exploration in this relatively new field.
Wenninger et al. [7] focused specifically on AI-assisted analysis for beach volleyball. Various models were evaluated, like Multilayered perceptron, convolutional neural networks (CNN), Recurrent neural networks - Gated recurrent unit RNN-GRU, and XGBoosted trees. The input variables were 3-dimensional Cartesian coordinates and two temporal coordinates and performance indicators (various metrics such as speed, accuracy, agility, strength, technique, strategy execution, etc..). The target for classification was the success of a rally, the attack direction, and the attack technique based on the events and/or positions that occurred in the rally before. The advances in this study suggest that the models performed better than random classification accuracy, ranging from 37 to nearly 60 percent for different tasks.

B. Use of Wearable Sensors

Wearable sensors, driven by advancements in sports science, offer unprecedented opportunities for biomechanical analyses outside the confines of a laboratory. Wang's exploration with micro inertial measurement units in volleyball, particularly assessing spike movements, shows this progression [8]. In this study, several ML classifiers were tested for accuracy using acceleration data. Comparing the classifiers showed that SVM achieves a high accuracy (94%) in assessing the volleyball spiking skill level. The results could help coaches and athletes keep track of condition changes during a training season.

This analysis methodology is further exemplified in basketball in a study with the overall objective of developing an advanced computational model to classify the skill level of basketball players during free throw shots using data from a single inertial sensor [9]. The results obtained, with classification accuracy, sensitivity, and specificity, were greater than 85 percent.

Traditional evaluations, reliant on extensive setups with motion-capture cameras, are giving way to more agile approaches utilizing wearable electronics. It is also worth mentioning that in our case, it was difficult and even impossible to use optical, IR-based motion-capture-based systems because volleyball was played outside.

C. Spectrograms for CNN Classification Tasks

Spectrograms, time-frequency representations of signals, are widely employed as input features for ML models learning an embedding of time-series signals. A common approach across multiple domains is to leverage the spectrogram as a visual representation of the time series and combine this with the use of a Convolutional Neural Network (CNN), an architecture that can learn a representation of the data's spatial structure. For example, spectrogram representations of audio signals are commonly combined with CNNs for tasks such as sound event detection [10] and soundtrack classification [11], as well as several natural language processing tasks [12]. Similarly, CNNs in combination with spectrogram representations of signals acquired from motion sensors have been applied in various sports and health science tasks, such as gait classification [13], human activity recognition [14], and sport activity classification [15]. Additionally, this was the approach taken by Guo et al. [9] in the basketball skill level classification task mentioned in the previous section, and which we likewise employ in the work presented in this paper.

D. Explainability

To have a comprehensive understanding of our methodology, the details of LIME have to be explored. LIME, which stands for Local Interpretable Model-agnostic Explanations, is a widely utilized tool in AI explainability research. Developed to enhance the interpretability of complex ML models, it operates by approximating the local linear behavior of a black-box model, making it model-agnostic meaning applicable to a range of classifiers.

Originating from the work of Marco Tulio Ribeiro and his collaborators, LIME aims to shed light on the decision-making processes of ML models. In our case, it is employed to discover the predictions of our Convolutional Neural Network model.

Our choice of using LIME stems from its effectiveness in generating human-understandable explanations for black-box models and the use of spectrograms. It constructs a simpler, interpretable model to approximate the complex decision boundaries of the primary model by perturbing input data and observing the model's response. This process enables us to visualize and understand the areas of the input space that influence the model's predictions.

LIME has demonstrated its utility beyond the realm of sports, finding application in various domains. For instance, it has been successfully employed in the National Basketball Association, as evidenced by the work of Wang et al. [16], showcasing its versatility and effectiveness in diverse contexts.

In conclusion, our unique contribution lies in applying LIME to spectrogram images from our CNN model. This not only expands the application of LIME within the sports domain but also contributes to the broader understanding of how explainability tools can resolve the complexity of ML models.

E. Summary

In summary, while traditional sports such as football and running have seen significant applications of AI and ML, less popular sports like beach volleyball are only beginning to explore these advancements. The use of wearable sensors, ML models like CNNs, and techniques like spectrograms are contributing to more nuanced and detailed sports analysis. Furthermore, the focus on explainability through tools like LIME is important in making these sophisticated models more accessible and interpretable for stakeholders.

III. Methodology

A. Dataset

Data was collected from 8 participants, comprising 4 world-class elite and 4 semi-elite players. For the world-class elites, two of these are the focus of the data collection and represent the highest level of elite play. The other two also represent players among the best, but not quite to the extent of the first two. This was collected as part of the collection of a wider dataset involving multiple motion and physiological sensors. We used acceleration data gathered from each participant through the upper torso in our scope. This was collected across an Equivital Lifemonitor worn in a vest[1]

The dataset, sourced from the Equivital device, encompasses:

- **Acceleration:** Sampled at 256Hz in milli-g along three axes.

The data extracted from the Equivital accelerometers is in the form of comma-separated values (CSV) files, where each line provides acceleration data along its lateral, longitudinal, and vertical axes.

After some early testing, we quickly discovered that the Equivital data was relatively stable both spatially and temporally and that it gave good initial results in classification.

Fig. 2: Equivital monitor with vest

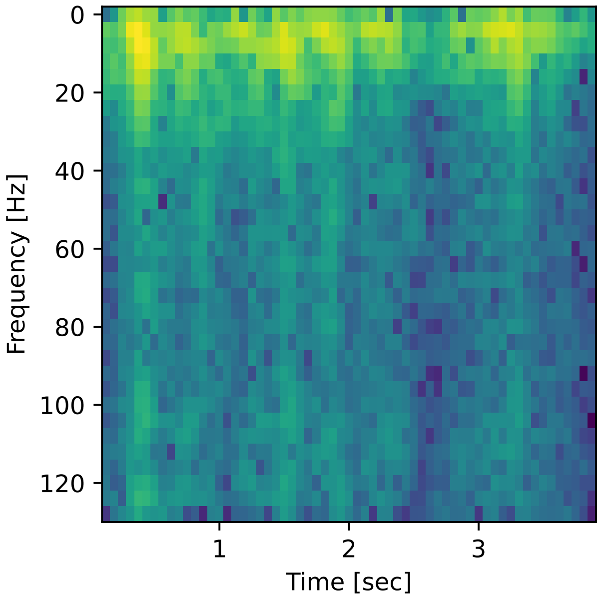
B. Data tagging

We employed an expert volleyball scout to code a video recording of the training session with the software Data Volley 4. This resulted in a series of timestamps for each player relating to the performance of the various plays in beach volleyball: serves (S), passes (R), sets (E), attacks (A), blocks (B), and defenses (D), along with their outcomes (e.g., effective, error, or neutral). We extracted x second segments from the acceleration data according to these timestamps

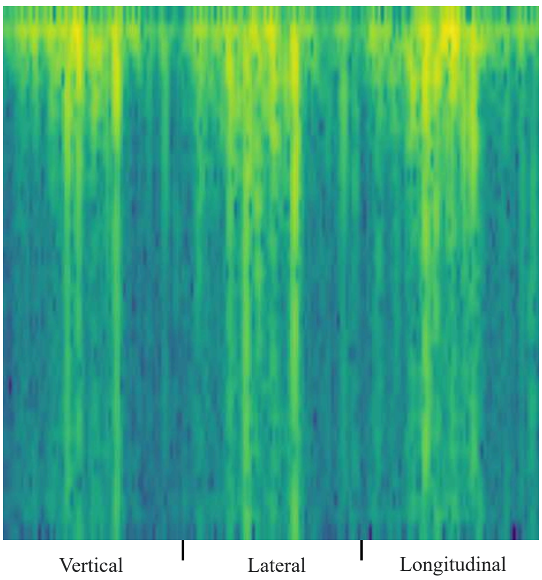
[1]`https://equivital.com/products/eq02-lifemonitor`

C. Models

1) *Input features*:
   Data extraction: We first collected the labels corresponding to the movements we were interested in (passes, sets, and attacks), and chose only the labels that represented successful attempts at the given moves.
   Then, to locate the specific lines of raw accelerometer data we wished to use, we cross-referenced the timestamps of the raw data with that of the collected labels. For each move we then extracted a total of 4 seconds of data, starting 2 seconds before and ending 2 seconds after the given label timestamp. These 4 seconds constitute one single sample. Considering the accelerometer frequency of 256Hz, this equates to 1024 lines of raw data per sample. To increase the total number of samples, we created three different samples from each successful movement attempt, each offset by 0.25 seconds or 64 lines. Augmenting the data in this way is a useful tool when dealing with smaller datasets [17]. The samples were then normalized individually. In total, we end up with 273 samples for passing moves, 317 samples for setting moves, and 480 samples for attacking moves. Samples of the separate moves are collected in three separate datasets, one set for each type of move.
   Data transformation: Next, using a Fourier transform, we transformed the data from the time domain to the frequency domain. This transformation enabled us to create the time-frequency spectrogram representations of each sample. We did this by utilizing the Hamming window function, a segment length of 64, and an overlap of 48. The Hamming window function was chosen for its good frequency resolution [18]. Considering that the data was recorded at 256Hz, the segment- and overlap values were chosen to give good time resolution.
   Spectrogram generation: Using these transformed samples, we now create separate spectrogram images for the three axes of acceleration, before concatenating the three into one single file (Figure 3). Then, we export the concatenated spectrograms as portable network graphics (PNG) image files and use them as inputs for the network, retaining spatial and temporal features from the original data. The final concatenated image consists of the outputs of the Fourier transforms side by side along the horizontal axis, while the vertical axis represents frequency ranges. The brighter the image is in a certain area, the stronger the corresponding frequency in the source data.
2) *Target variables* :
   Due to the importance of their role in scoring in beach



(a) Spectrogram for a single axis of acceleration. For illustration purposes, this image has added information about frequency and time.



(b) Concatenated spectrogram of all three axes of acceleration. For illustration purposes, the divides have been highlighted and labeled.

Fig. 3: Spectrograms

volleyball, we decided to focus on passing, setting, and attacking movements. This includes analyzing to what degree a given movement (e.g. an attack) is performed differently by the players at the two skill levels, and also dissecting what these differences are. The network thus had to find patterns and features in the spectrogram images that signify they originated from a player at either an elite or a semi-elite level. The targets were the player numbers as coded by the scout/groupings of the players into elite/semi-elite.

3) *Convolutional Neural Network (CNN)*:
In this project, we opted for the use of CNNs because they are widely recognized in the field, especially for image classification tasks [9]. They offer a straightforward and intuitive approach to processing our spectrograms, which are represented as images. Their ability to identify intricate patterns in images aligns well with our goal of understanding the complex spatial aspects of beach volleyball performance. Spectrograms don't include any spatial information, however, we did construct spectrogram inputs in which spatial information was contained in the form of concatenating the three axes (vertical position relates to frequency, and horizontal position relates to both time and acceleration).

4) *Training*:
The CNN (Figure 4) was trained on a series of spectrogram images using a 90-10 validation split. It uses a Tensorflow sequential model with two 2D convolution layers, both with Rectified Linear Unit (reLU) activation functions. The first layer has 32 kernels while the second has 64, with respective sizes of sizes of 5x5 and 2x2 respectively. The data is max-pooled with a size of 2x2 after each activation and passes through one "flatten" and one "dense" layer at the end before the final sigmoid output function. There are also dropout layers after each max-pool layer which reset 20 percent of the weights to combat overfitting. We chose the Adam optimizer for its superior performance, especially when working on noisy data [19].
We then trained three separate models, one for each of the three datasets (passes, sets, and attacks), with slightly different hyperparameters. The models were each trained for 20 epochs with a batch size of 32. The loss was calculated using binary cross-entropy. For the passing and attacking samples, we used a learning rate of 0.0002. For the setting samples, we used a learning rate of 0.0001.

5) *Local Interpretable Model-Agnostic Explanations (LIME):*
LIME is then used to highlight the positive regions in the correctly predicted spectrograms from our CNN model for each distinct move type.

6) *Evaluation* :
The evaluation of our model is centered on the metrics of accuracy, precision, recall, and the F1 score.
However, our data posed a significant challenge as it was difficult to isolate the specific features we wanted our CNN model to focus on. An example of this is that a given player could have attributes in their movement pattern that are not necessarily correlated to their skill level. These unwanted features could then help the models classify the players correctly on the wrong grounds.
To address this issue, we created test sets with data exclusively obtained by players that were not included in the training data. This means, for example, that we

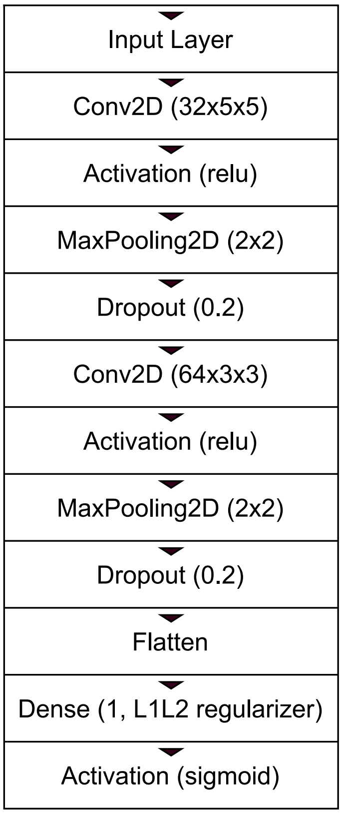


Fig. 4: Layers of our CNN model

would train on data from players 1, 2, and 3 while testing on data exclusively from player 4. In this way, we can evaluate model performance purely on the task of classifying different skill levels, isolated from any individual player characteristics.

The test sets therefore vary somewhat in size, depending on the number of available samples. For passing moves, the test set had a size of 63 (36/27) which gave us a 77-23 train-test split. For setting moves, the test set had a size of 105 (51/54) which gave us a 67-33 train-test split. Finally, for attacking moves, the test set had a size of 102 (51/51) which gave us a 79-21 train-test split.

An additional challenge is that the elite and semi-elite games were played on different days under slightly different conditions. Therefore, there could be differences in the background noise of our data that can help our model separate the two different skill levels. Higher levels of moisture in the sand on one day could for example alter the deceleration of a player landing from a jump, compared to dryer sand.

To account for these irrelevant differences, we created additional test sets. In these sets (hereafter referred to as control sets) however, the samples are still extracted from the players in the normal test sets, but from random timestamps. In other words, the control sets contain only noise that is randomly selected from the dataset. If there is nothing in the background data that the model can use in training, we would expect no better accuracy than 50 percent from the control sets, similar to random guessing. Any performance above this would indicate that the model uses unwanted background factors to separate elites from semi-elites, and must then be taken into account.

IV. Results

A. Discriminating elites and semi-elites

The results of our trained models in discriminating elite from semi-elite players are shown in Table I, Table II, Table III, and Table IV. More precisely, these results show the mean performance of each model when trained and tested with unchanged hyperparameters 50 consecutive times. They also show the corresponding confidence intervals (CI) at 95% and the standard deviations (SD).

TABLE I: Accuracy scores from the test sets

	Mean	CI	SD
Pass	0.826	± 0.008	0.027
Set	0.729	± 0.034	0.119
Attack	0.865	± 0.024	0.082

TABLE II: Recall scores from the test sets

	Mean	CI	SD
Pass	0.836	± 0.021	0.072
Set	0.964	± 0.018	0.062
Attack	0.987	± 0.034	0.010

TABLE III: Precision scores from the test sets

	Mean	CI	SD
Pass	0.863	± 0.014	0.050
Set	0.678	± 0.039	0.134
Attack	0.809	± 0.028	0.097

Table IV and Table V show the results of our models evaluated on the test and control sets. These scores originate from the same 50 consecutive runs as the ones above. We use the F1 score because it equally weights both false positives and false negatives. This is advantageous in cases where the sample sizes are slightly unbalanced. These results highlight the difference in performance when evaluating our trained models on the test data vs the control data (see details under subsection III-C "Evaluation"), with confidence intervals at 95%.

TABLE IV: F1 scores from the test sets

	Mean	CI	SD
Pass	0.845	± 0.008	0.028
Set	0.784	± 0.020	0.071
Attack	0.885	± 0.017	0.058

TABLE V: F1 scores from the control sets

	Mean	CI	SD
Pass	0.502	± 0.032	0.113
Set	0.602	± 0.022	0.078
Attack	0.707	± 0.039	0.011

For the passes, sets, and attacks in the test sets, the Test F1 scores are as you would expect from observing the precision and recall. Regarding the control F1 scores, we can see that it is close to 0.50 for the passing moves. This means that, when classifying in this category, the model is not finding any features in the background noise. Rather, it is finding features that most likely originate from the respective body movement.

As for the setting moves, the control F1 score is somewhat higher. This might indicate that the model is finding features in the background noise and using them to classify with slightly better performance than random guessing.

When it comes to attacking movements, we can see that the control F1 score is even higher. This means that the models were able to classify control samples with higher performance than random guessing, indicating that there were factors other than the attacking moves that it picked up on. This might also explain why the models trained on attacking moves perform the best across the board. However, since there is still a gap of 0.178 between the means of the test F1 and control F1 scores of the attacking moves, the model is also able to find patterns that most likely originate from the attacking moves to some degree.

For all three pairs of test and control F1 scores, Mann-Whitney U rank tests were completed with the null hypothesis being that the two sample distributions are equal. We chose this test based on a histogram analysis that concluded the data does not follow any normal distribution. For all three pairs, the tests achieved values of $p < 0.001$, strongly indicating statistically significant differences between test and control performance, allowing us to discard our null hypothesis.

B. LIME-analysis

Figure 5 shows an example of the spectrogram data our model sees (the "Original Image") above the type of explanation we can expect from LIME (the "Explanation Image"), with the red and yellow lines highlighting the regions that played a positive role in the CNN model's prediction of an elite player for this input data (which represents an elite player's attack). By interpreting the Explanation Image we can begin to understand why the CNN model classified the Original Image as an elite player move. For this example, we see that the LIME explanation emphasizes specific movements, mainly lateral ones ranging from 0.1 Hz to approximately 115 Hz, with additional smaller regions in longitudinal movements featuring frequencies between 0.1 Hz and 80 Hz. After collecting Explanation Images for each move type—passes, sets, and attacks, all generated by the same trained model, we employed an aggregation process. Aggregation, in this context, involves layering each Explanation Image on top of the others, creating a composite view. This method allows us to unveil shared regions and frequencies crucial for accurate predictions from our CNN and LIME models across various move types. It's essential to clarify that these Aggregated Explanation Images are derived from different instances of the same trained model, each corresponding to a specific move type. The resulting Aggregated Explanation Images, illustrated in Figure 6 to Figure 8, further emphasize this, with lighter regions signifying higher importance and darker regions vice versa.

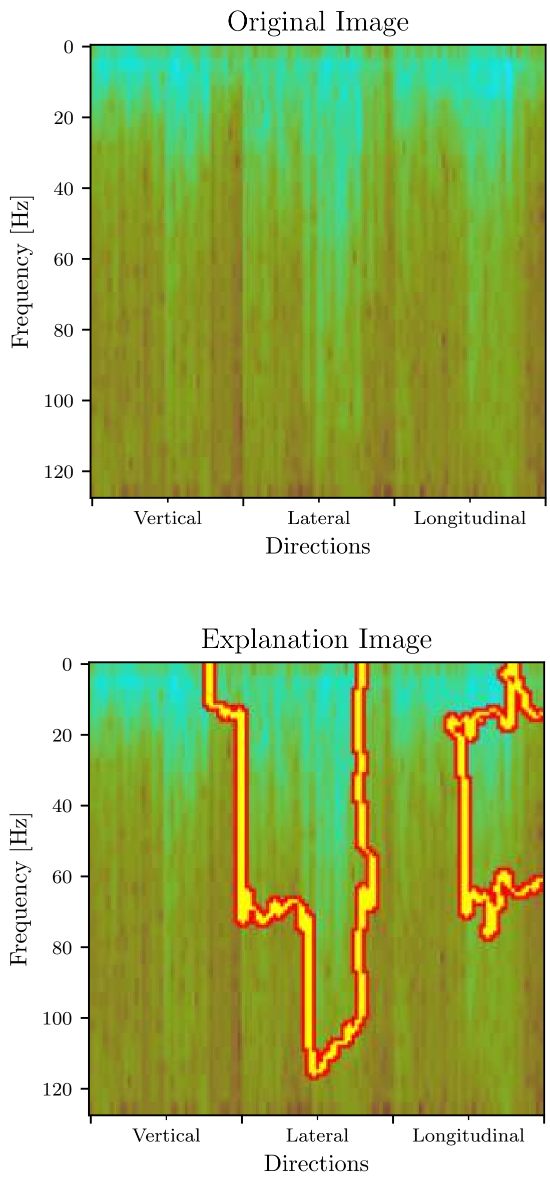


Fig. 5: Explanation Of An Elite-player Attack Move From LIME

Our analysis of the Aggregated Explanation Mask images for elite players' passes (shown in Figure 6) reveals that the

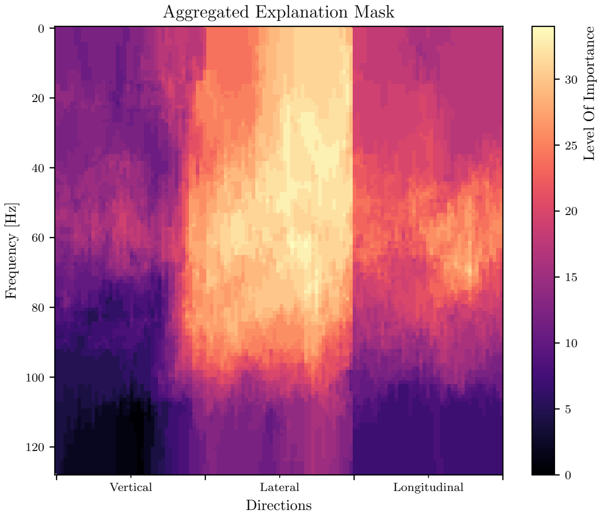


Fig. 6: Aggregated Explanation Mask For Elite Passes

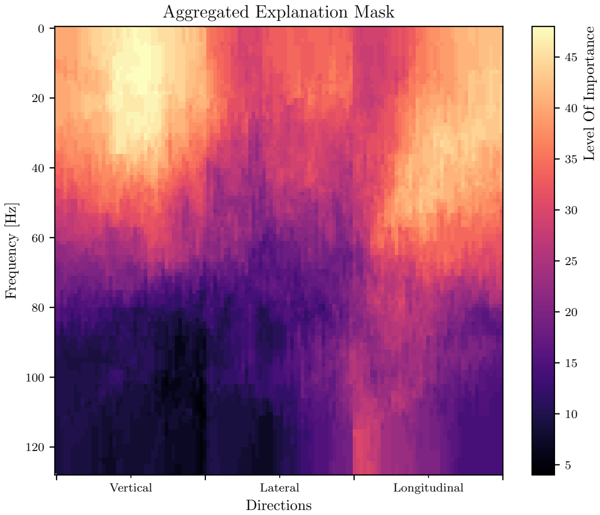


Fig. 7: Aggregated Explanation Mask For Elite Sets

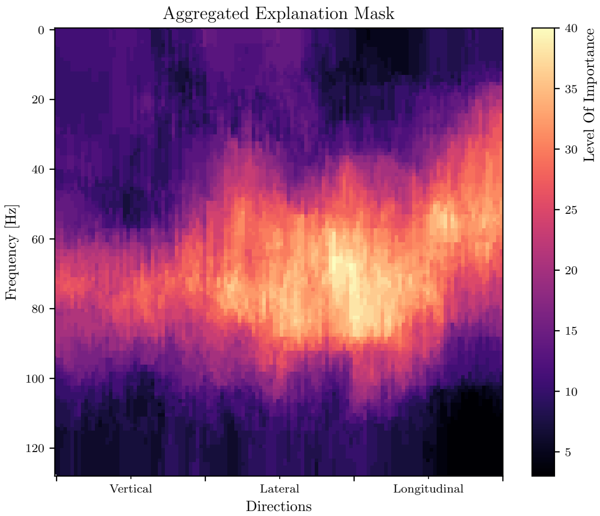


Fig. 8: Aggregated Explanation Mask For Elite Attacks

focus of our CNN model's prediction is mainly on lateral movements with frequencies ranging from 0.1 Hz to about 90 Hz, with more focus in the 20 Hz to 70 Hz range. In the context of beach volleyball, lateral movements are crucial for successful receiving actions, as players need to quickly adjust their position to the trajectory of the ball. Similarly, for set moves (as shown in Figure 7), the focus is on vertical and longitudinal movements with frequencies ranging from 0.1 Hz to roughly 65 Hz. This aligns with the strategic importance of precise vertical and longitudinal movements in setting up plays. More emphasis on vertical movements of frequencies 0.1 Hz to 35 Hz indicates the specific nuances associated with setting actions in beach volleyball.

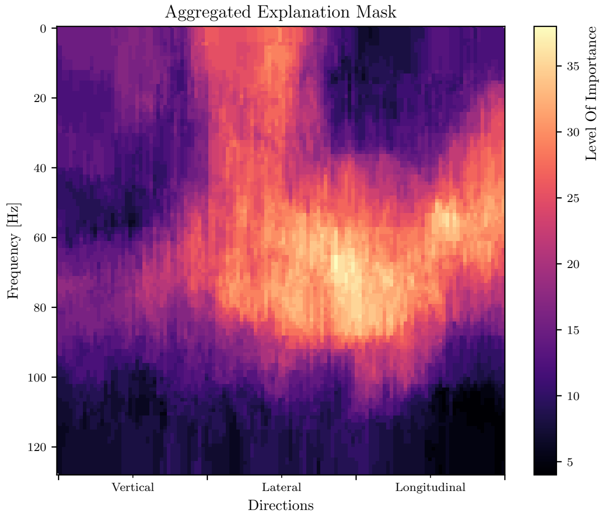
Our findings remained consistent across various trials of the Aggregated Explanation Images for attacks, as depicted in Figure 9. This emphasizes the crucial role of specific frequency ranges (15 to 85 Hz) in lateral and longitudinal movements for distinguishing between player classes. When we say 'multiple runs', we mean different instances of the same trained model, each associated with different accuracies observed during the testing phase of the CNN model. This sustained consistency can be attributed to maintaining constant hyper-parameters, ensuring the reproducibility of LIME plots. The uniformity observed in these diverse runs underscores the model's resilience in recognizing key features, providing reliable insights into player classification across various scenarios.

V. Discussion

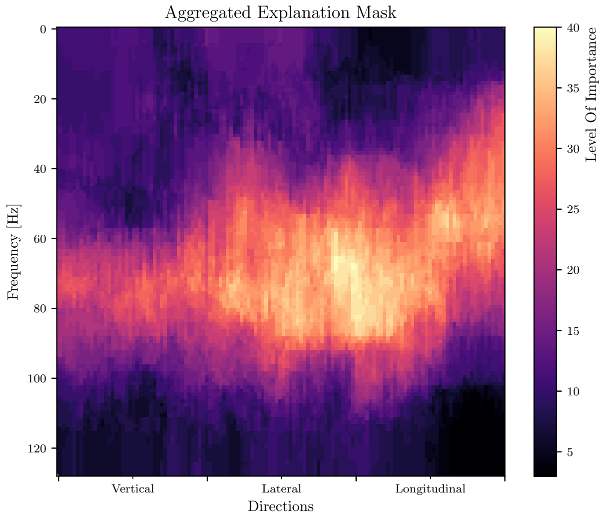
These initial results are promising, with the passes being particularly auspicious. As previously stated, these results are obtained from predicting the class of samples from players not included in the training sets. Therefore, even with the limited size of our dataset, they should not be severely impacted by overfitting. They should also not be affected by recognizable elements of a given player's movement pattern.

One of the main challenges of this project was to find a robust methodology with which to determine the success of our classifiers. We ultimately decided to utilize the test and control sets and compare the respective scores with one another. Taking these comparisons into consideration, we see a statistically significant performance increase in our test data versus our control data. This means that the classifiers are in fact able to recognize differences in athlete movements related to their skill level. When we also consider the relatively small differences in skill level between the elite and semi-elite players to an average player, our contributions assume additional value.

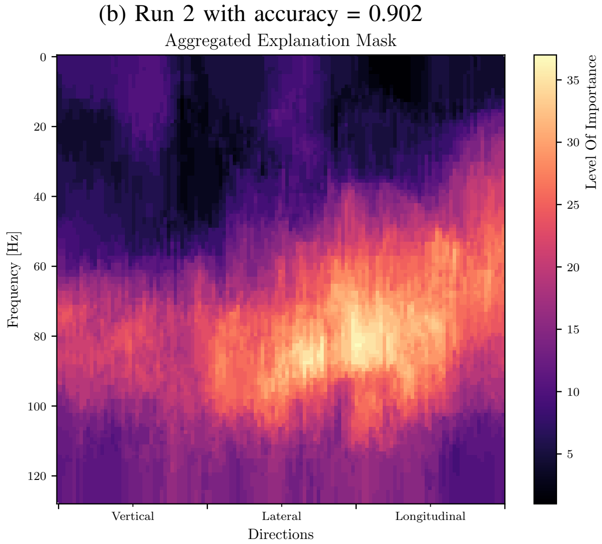
The high accuracy rates achieved in distinguishing between elite and semi-elite players underscore the potential of these methods in identifying nuanced differences in playing styles and techniques. Particularly noteworthy is the capability of our model to identify specific frequency ranges and movement patterns that are characteristic of elite players. These findings resonate with the broader goal of sports analytics to offer precise, data-driven insights that could



(a) Run 1 with accuracy = 0.85



(b) Run 2 with accuracy = 0.902



(c) Run 3 with accuracy = 0.823

Fig. 9: Multiple Runs Of The Aggregated Mask Images For Attack Moves

revolutionize training and performance evaluation in beach volleyball.

Through the interpretability provided by LIME's explainable images, our analysis uncovered consistent movement patterns across different volleyball actions—passes, sets, and attacks. While the challenge remains in transforming these abstract data patterns into actionable training interventions, these findings provide a foundational entry point for further investigation. Integrating this data-driven approach with video analysis could offer a more holistic view. By aligning the frequency importance from our model with video footage of player movements, coaches can visually correlate the model's insights with actual gameplay, making the data more tangible and actionable. This method facilitates a deeper understanding of how specific movements translate into successful gameplay strategies.

In dissecting these data patterns, it is observed that elite player passes are predominantly characterized by lateral movements spanning frequencies of 20 Hz to 70 Hz. These lateral movements, crucial during elite player passes, involve dynamic sideways motions that reflect the player's ability to optimally position themselves for effective ball control. Similarly, set moves primarily involve vertical and longitudinal movements, with a notable emphasis on vertical movements ranging from 0.1 Hz to 35 Hz. For attack moves, the distinguishing characteristics are found within frequency ranges of 15 to 85 Hz, encompassing both lateral and longitudinal movements.

These insights strongly suggest that certain movements within specific frequency ranges are pivotal for the accurate classification of player performances in beach volleyball. They provide valuable insights that could inform player development and coaching strategies, potentially enhancing training regimens and tactical approaches. [20].

To bridge the gap between frequency-based explanations and practical training applications, we propose developing a comprehensive framework that maps these spectrogram features to specific volleyball techniques and exercises. This framework could fro example include:

- **Lateral movement drills**: Based on the frequency range identified for elite passes, incorporate agility ladder drills, side-to-side shuffles, and reactive lateral movement exercises to improve players' quickness and positioning.
- **Vertical precision exercises**: For setting actions, focus on plyometric exercises, vertical jump training, and precise ball-handling drills to enhance vertical movements within the identified frequency range.
- **Attack power training**: Emphasize strength and conditioning exercises that enhance power generation in the identified frequency range for attacks, such as resistance band exercises, medicine ball throws, and plyometric push-ups.

From an applied perspective, the findings from our LIME analysis can serve as an entry point for further investigation into the critical aspects of elite beach volleyball performance. For example, the identified movement patterns and frequency ranges can be used to guide more detailed analyses of specific game situations and player actions. This approach can help coaches and analysts prioritize their focus, saving time and effort by concentrating on the most impactful aspects of player performance. In elite sports, where time and resources are limited, leveraging machine learning to identify key performance indicators can streamline the analytical process. By pinpointing the critical movements and frequencies that differentiate elite players, teams can allocate their resources more efficiently, focusing on refining these essential skills.

There are several limitations to this study. The constrained dataset size and the potential for overfitting necessitate a cautious interpretation of our results. Future studies could benefit from larger datasets, possibly encompassing a broader range of skill levels and more varied playing conditions. This could help in refining the model further and enhancing its applicability in an applied context.

Therefore, we advocate for further investigation in this area, ideally with a broader participant base to enhance the validity and applicability of the findings.

VI. Conclusion

By building on data from some of the world's best beach volleyball players, this study has given a unique insight into the possibility of using ML to discern elite from semi-elite players and explain what features of their motions are most important in this classification
Our findings have several implications. Firstly, the high accuracy in classifying player movements underscores the potential of ML in enhancing training and performance strategies. Coaches can leverage these insights to tailor training programs, focusing on specific movement patterns and frequencies characteristic of elite performance. This approach could lead to more effective training methodologies, potentially improving some standards of play in beach volleyball.

Furthermore, the application of LIME has brought an essential element of transparency and explainability to our model, providing coaches, players, and stakeholders with a view of the factors contributing to elite performance. This understanding is crucial for the ethical and responsible application of AI in sports, ensuring that decisions based on these models are well-informed and fair. However, while LIME has undoubtedly enhanced the transparency and explainability of our model, it is imperative to acknowledge its limitations. The interpretability granted by LIME primarily extends to local contexts around specific predictions, and extrapolating global model behavior solely from LIME explanations may lead to oversights. Thus, while LIME contributes significantly to transparency, there remain aspects beyond its scope, and caution should be exercised in drawing broader conclusions solely from LIME plots. This nuanced perspective is vital for a comprehensive evaluation of our model's performance and a responsible integration of AI in sports analysis.

It is important to acknowledge the limitations of our study concerning the size and diversity of the dataset. Future research could expand upon this work by incorporating a larger and more varied dataset, possibly including more nuanced player data and different levels of skill. Such expansion could enhance the model's accuracy and reliability, making it a more robust tool for player analysis.

In addition, the torso acceleration data employed in the current work comprises a small part of a larger dataset collected from the volleyball players, which also consists of acceleration data collected from sensors mounted on additional body parts such as the dominant wrist, as well as various physiological signals (ECG, respiration rate), and high-resolution video recordings. Leveraging the additional data modalities for multimodal learning approaches offers a further path for future research.

In summary, our study contributes to the growing body of knowledge in sports analytics by offering novel insights into the physical characteristics that differentiate elite beach volleyball players. It underscores the potential of ML in transforming sports training and strategy, providing a possible path for further research.

VII. Ethics statement

1) **Consent:** Prior to the data collection phase, informed consent was obtained from all the participants involved in the study. They were comprehensively briefed on the aims of the research, the methodologies employed, and the potential outcomes. All participants were ensured of their right to withdraw from the study at any given time without any consequences.
2) **Anonymity and Confidentiality:** The data obtained from the participants has been treated in line with GDPR. Personal identifiers were stripped from the dataset to ensure the anonymity of the participants. Our analysis does not focus on individual performances, but rather on general patterns that differentiate skill levels. Thus, specific identities linked to the data cannot be deduced from our findings.
3) **Data Handling and Storage:** Data obtained from the players, especially sensitive information such as ECG and respiration rates, have been securely stored in encrypted formats. Access to this data is restricted to the primary researchers of this project. Upon the conclusion of this research, all raw data will be stored securely for a stipulated duration, post which it will be responsibly disposed of.
4) **Transparency:** All methodologies and processes applied in this study have been transparently communicated in the paper. This includes not only the data collection methods but also the algorithms and analysis techniques employed.
5) **Potential Implications:** We recognize the implications of our findings, especially for athletes and trainers who might consider integrating ML tools in their training regimes. While our results aim to provide insights into movement patterns, they are not definitive judgments of players' abilities. As such, they should be

interpreted with caution and supplemented with human expertise.

6) **Helpfulness:** The primary intention behind this research is the advancement of knowledge in the fields of sports analysis and ML. We believe our findings can contribute positively to training methodologies and enhance the sport of beach volleyball. However, we also advise caution in directly implementing any recommendations without considering the broader context and individual differences.
7) **Explainability:** Explainability is a key ethical consideration, especially in deep learning applications like ours, which are often perceived as black boxes. By incorporating LIME to introduce an explainable component, we enhance the transparency and trustworthiness of our CNN-based model. This move towards greater explainability allows stakeholders to understand, trust, and critically evaluate the model's outputs, which is essential for ethical AI deployment in sports. It not only facilitates easier identification and correction of potential errors but also contributes to accountability and fairness by exposing any underlying biases in the model. Consequently, this approach ensures informed decision-making by providing clear insights into how the model processes and analyzes data. As sports analytics increasingly influence critical decisions in training and strategy, our commitment to explainable AI aligns with the ethical standards of transparency and fairness, crucial for maintaining the integrity of the sport and its athletes.

VIII. Acknowledgements

This work is supported by the Research Council of Norway as a part of the projects: Collaboration on Intelligent Machines (COINMAC) project, under grant agreement no. 309869, Predictive and Intuitive Robot Companion (PIRC) under grant agreement no. 312333 and through its Centres of Excellence scheme, project number: 262762.

References

[1] J. W. Roger Bartlett and M. Robins, "Is movement variability important for sports biomechanists?" *Sports Biomechanics*, vol. 6, no. 2, pp. 224–243, 2007, pMID: 17892098. [Online]. Available: https://doi.org/10.1080/14763140701322994

[2] L. Choo, A. Novak, F. M. Impellizzeri, C. Porter, and J. Fransen, "Skill acquisition interventions for the learning of sports-related skills: A scoping review of randomised controlled trials," *Psychology of Sport and Exercise*, vol. 72, p. 102615, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1469029224000268

[3] A. Rajšp and I. Fister, "A systematic literature review of intelligent data analysis methods for smart sport training," *Applied Sciences*, vol. 10, no. 9, p. 3013, 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/9/3013

[4] O. M. H. Rindal, T. M. Seeberg, J. Tjønnås, P. Haugnes, and Sandbakk, "Automatic classification of sub-techniques in classical cross-country skiing using a machine learning algorithm on micro-sensor data," *Sensors*, vol. 18, no. 1, 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/1/75

[5] M. Z. Uddin, T. M. Seeberg, J. Kocbach, A. E. Liverud, V. Gonzalez, Sandbakk, and F. Meyer, "Estimation of mechanical power output employing deep learning on inertial measurement data in roller ski skating," *Sensors*, vol. 21, no. 19, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/19/6500

[6] R. Derie, P. Robberechts, P. Van den Berghe, J. Gerlo, D. De Clercq, V. Segers, and J. Davis, "Tibial acceleration-based prediction of maximal vertical loading rate during overground running: A machine learning approach," *Frontiers in Bioengineering and Biotechnology*, vol. 8, 2020. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fbioe.2020.00033

[7] S. Wenninger, D. Link, and M. Lames, "Performance of machine learning models in application to beach volleyball data," *International Journal of Computer Science in Sport*, vol. 19, no. 1, pp. 24–36, 2020. [Online]. Available: https://sciendo.com/article/10.2478/ijcss-2020-0002

[8] Y. Wang, Y. Zhao, R. H. Chan, and W. J. Li, "Volleyball skill assessment using a single wearable micro inertial measurement unit at wrist," *IEEE Access*, vol. 6, pp. 13 758–13 765, 2018. [Online]. Available: https://scholars.cityu.edu.hk/files/119000989/21930976.pdf

[9] X. Guo, E. Brown, P. Chan, R. Chan, and R. Cheung, "Skill level classification in basketball free-throws using a single inertial sensor," *Applied Sciences*, vol. 13, no. 9, p. 5401, 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/9/5401

[10] F. Demir, D. A. Abdullah, and A. Sengur, "A new deep cnn model for environmental sound classification," *IEEE Access*, vol. 8, pp. 66 529–66 537, 2020.

[11] S. Hershey, S. Chaudhuri, D. P. W. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, M. Slaney, R. J. Weiss, and K. Wilson, "Cnn architectures for large-scale audio classification," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 131–135.

[12] T. Arias-Vergara, P. Klumpp, J. C. Vasquez-Correa, E. Nöth, J. R. Orozco-Arroyave, and M. Schuster, "Multi-channel spectrograms for speech processing applications using deep learning methods," *Pattern Analysis and Applications*, vol. 24, no. 2, p. 423–431, May 2021.

[13] M. D. Nguyen, K.-R. Mun, D. Jung, J. Han, M. Park, J. Kim, and J. Kim, "Imu-based spectrogram approach with deep convolutional neural networks for gait classification," in *2020 IEEE International Conference on Consumer Electronics (ICCE)*, 2020, pp. 1–6.

[14] C. Ito, X. Cao, M. Shuzo, and E. Maeda, "Application of cnn for human activity recognition with fft spectrogram of acceleration and gyro sensors," in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, ser. UbiComp '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1503–1510. [Online]. Available: https://doi.org/10.1145/3267305.3267517

[15] Y.-L. Hsu, H.-C. Chang, and Y.-J. Chiu, "Wearable sport activity classification based on deep convolutional neural network," *IEEE Access*, vol. 7, pp. 170 199–170 212, 2019.

[16] Y. Wang, W. Liu, and X. Liu, "Explainable ai techniques with application to nba gameplay prediction," *Neurocomputing*, vol. 483, pp. 59–71, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231222001333

[17] C. Shorten and T. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Big Data 6*, vol. 6, p. 60, 07 2019. [Online]. Available: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0197-0

[18] P. Podder, T. Khan, M. Khan, and M. Rahman, "Comparative performance analysis of hamming, hanning and blackman window," *International Journal of Computer Applications*, vol. 96, pp. 1–7, 06 2014. [Online]. Available: https://www.researchgate.net/publication/271156675_Comparative_Performance_Analysis_of_Hamming_Hanning_and_Blackman_Window

[19] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014. [Online]. Available: https://arxiv.org/abs/1412.6980

[20] T. Kautz, B. H. Groh, J. Hannink, U. Jensen, H. Strubberg, and B. Eskofier, "Activity recognition in beach volleyball using a deep convolutional neural network," *Data Mining and Knowledge Discovery*, vol. 31, pp. 1678–1705, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:16170944