



# **Explaining Cricket Shot Techniques with Explainable AI**

**A deep dive into the possibilities of XAI implemented on pose-estimation based  
cricket shot classification**

---

**Bruno Martinović**  
TU Delft

A Thesis Submitted to EEMCS Faculty Delft University of Technology,  
In Partial Fulfilment of the Requirements  
For the Bachelor of Computer Science and Engineering  
June 22, 2025

Name of the student: Bruno Martinović  
Final project course: CSE3000 Research Project  
Thesis committee: Ujwal Gadiraju, Danning Zhan, Mark Neerinx

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.

Submitted to CSE TU Delft

## Abstract

Recent advancements in pose estimation, activity classification, and explainable artificial intelligence (XAI) have opened new opportunities in sports analytics. However, their combined application within the domain of cricket remains largely unexplored. This paper investigates the integration of XAI methods to interpret black-box models trained to classify cricket shot techniques using pose keypoints extracted from video data. I implement and compare several techniques, including SHAP, Grouped SHAP, Feature Importance, Permutation Importance, LIME, Grad-CAM, and Accumulated Local Effects (ALE) plots based on their ability to generate meaningful and interpretable explanations. Through a structured experimental pipeline involving MediaPipe-based keypoint extraction, random forest (RF), convolutional neural networks (CNNs), and multiple explanation methods, I evaluate which techniques most effectively highlight the body keypoints critical to accurate shot classification. The findings indicate that SHAP outperforms other methods due to its ability to generate both local and global explanations, along with intuitive visualizations. This work contributes to the development of transparent sports AI systems and lays the foundation for future applications of interpretable machine learning in athletic coaching and skill assessment.

## 1 Introduction

With the rapidly evolving AI technologies, more and more are becoming incorporated into sports, be it in the form of AI-assisted VAR, game predictions, or technique checking. AI is quickly becoming a must-have tool for athletes looking to improve in their respective sports, and one of those sports is cricket. Cricket in itself has many aspects, but one of the most difficult to master is shot technique. AI can be used to help athletes and coaches analyze shot techniques; however, the black box nature of many AI-driven pose estimation models limits their practical utility, particularly in a complex and technique-driven sport like cricket.

Implementing explainable AI into the shot analysis pipeline would be helpful on all levels of the sport, not only for improving one's technique but also for preventing injury, because with better technique, there is a lower risk of injury. At the beginner level, a hobbyist player could analyze their shots and get better at the sport without having to make a commitment, such as hiring a coach. At the amateur and pro level, a player wishing for more training is able to analyze their own shots and refine their technique. Coaches could also use this tool as an assistant coach to help them get a different perspective and spot elements they might have missed to improve their players. Overall, there are lots of benefits from creating this tool.

### 1.1 Pose Estimation and Explainable AI in Sports

Pose estimation has become a powerful tool in sports analytics, enabling the extraction of structured representations of human movement for tasks such as action recognition, technique evaluation, and injury risk assessment. Applications in sports like tennis by Kurose et al. [2018], and soccer by Mauricio Salazar and Alatrística-Salas [2024] have demonstrated the value of 2D and 3D keypoint tracking in improving both coaching insights and model performance, as well as technique classification. Despite this progress, interpreting and explaining the predictions of models that use pose estimation results, particularly black-box models, remains a critical challenge that has not yet been deeply explored.

### 1.2 Research in Pose Estimation with XAI

Recent advancements in explainable AI (XAI) have enabled better interpretation of models that use pose estimation as input. Qiu et al. [2024] introduced XPose, an explainable pose estimation framework that uses Grouped Shapley Values to enhance the interpretability of keypoint importance. This is an extension of SHAP (SHapley Additive exPlanations), a unified framework for interpreting predictions by attributing contributions to each feature, originally proposed by Lundberg and Lee [2017].

Other XAI approaches have also been employed in the context of human pose and image data classification explanation. Grad-CAM, introduced by Selvaraju et al. [2020], generates heatmaps by computing gradients of the output with respect to convolutional layer activations, and has been widely used to interpret CNN-based pose models. Dindorf et al. [2021] applied LIME (Local Interpretable Model-Agnostic Explanations) to posture classification in the medical domain, providing local, human-understandable explanations by perturbing input samples and fitting an interpretable surrogate model. Ribeiro et al. [2016], who originally proposed LIME, emphasize its model-agnostic nature and its applicability across tabular and image data. Additionally, RISE, proposed by Petsiuk et al. [2018], introduces a black-box explanation method based on randomized masking and input sampling, yielding heatmaps even for models that are otherwise opaque. Each of these techniques varies in granularity and applicability. SHAP and LIME work well with structured input, while Grad-CAM and RISE are suited for CNNs processing image-like data.

Despite these advances, current work predominantly focuses on general pose estimation and classification, or medical use cases. The integration of XAI into domain-specific pose-based classification, such as sports techniques, is still in its early stages.

### 1.3 Research in Cricket Performance Analysis

In the domain of cricket, machine learning has been used to analyze various aspects of player performance, including shot classification and gesture recognition. Ahmad et al. [2023] proposed a deep learning-based architecture combining CNN and LSTM to classify cricket shot techniques from video input. A different, but also viable approach was taken by Siddiqui et al. [2023] where MediaPipe was employed for pose extraction and compared multiple machine learning models for cricket shot prediction based on extracted keypoint features. Datta et al. [2024] followed a comparable approach using MoveNet for pose estimation and a neural network classifier, reflecting the growing trend of using black-box models in cricket analytics. Earlier work by Kumar et al. [2014] used optical flow analysis to classify stroke direction based on movement dynamics rather than pose. Meanwhile, Mili et al. [2022] explored a different use case within cricket: umpire gesture recognition, which they approached using MobileThinNet and an artificial neural network (ANN) for classification of five standardized signals.

### 1.4 Gap in Literature

While research has separately explored XAI for pose estimation and general classification, as well as the use of machine learning for cricket performance analysis, **no study to date has integrated XAI into pose-based cricket shot classification models**. This presents a significant opportunity to bridge interpretability with predictive accuracy. By applying explainable AI techniques to keypoint-based and CNN-based cricket shot classification models, one can identify which joints or movements are most influential for particular shot types. Such insights could support data-driven coaching, improve model trust, and enable fairer, more transparent sports analytics systems.

**How can Explainable AI methods simplify the understanding of pose-based classification results for cricket shots?** is my research question. The focus of my research is not to develop new XAI techniques, but to explore how existing ones can be effectively applied to interpret pose estimation data in the context of cricket performance analysis. This includes identifying which body key points are most relevant for understanding various cricket shots, how they can be visualized meaningfully, and examining the interdependencies between them.

The research question can therefore be split into two subquestions:

1. Which lightweight XAI methods (e.g., Grad-CAM or SHAP) can provide meaningful insights for pose-based and CNN-based classification models?
2. How can XAI techniques highlight the contributions of individual keypoints and their interdependencies in cricket poses?

Through this paper, I determine which XAI techniques are most effective in this domain, identify the most informative key points to visualize, and clarify how these key points relate to one another in the execution of different cricket shots.

The Methodology section outlines the approach taken to identify the most appropriate XAI techniques for pose estimation in the context of cricket. In the experimental analysis section, a description of the pipelines that have been implemented is given. The Results section presents an evaluation of these techniques, emphasizing the most influential keypoints for shot classification and their interdependencies, and provides an example shot evaluation. The Limitations section goes over the problems encountered during the project. The Responsible Research section reflects on the study's ethical implications and assesses the methodological framework's reproducibility. Lastly, the Discussion, Conclusion, and future works sections compare the findings with related research and highlight the XAI techniques deemed most effective for interpreting cricket shot mechanics, as well as propose further research that can be performed in the field.

## 2 Methodology

This research aims to determine which XAI methods can be used to provide meaningful insight into cricket shot analysis and finding weight of contributions of body keypoints (e.g. left shoulder, right elbow) and combinations of such. This will be achieved by performing a literature search into different XAI models applicable to classification models and CNNs, finding a dataset of cricket shots, selecting the most promising and viable XAI models, implementing them, and finally evaluating their usefulness for the purpose of cricket shot analysis.

### 2.1 Selected Pose Estimation and Classification models

Through the literature search, multiple viable methods were found for cricket shot classification, the two taken as inspiration for this project were:

Enhancing Cricket Performance Analysis with Human Pose Estimation and Machine Learning Siddiqui et al. [2023] - In this paper, multiple classification models are compared when paired with MediaPipe, a pose estimation model which provides body keypoints split into x,y,z coordinates when it is run on an image/frame. The classifiers are then trained and evaluated on these keypoints. The classification models compared are Long Short-Term Memory (LSTM), Logistic Regression, Decision Tree, k-nearest neighbor, support vector machine, and Random Forest (RF). The conclusion of this paper is that the RF classifier performed best, boasting a 0.998 precision, recall and F1 score.

Optimized deep learning-based cricket activity focused network and medium scale benchmark Ahmad et al. [2023] - This paper provides a dataset that contains 722 cricket shot videos. The dataset contains different classes of batting techniques, including pull shot, bowled, reverse-sweep, defence, and cover drive. The videos are between 2-3 seconds long, and most are of the same format, meaning the camera shows the different people in the video at approximately the same position every time. This paper also proposes an end-to-end deep learning model for shot classification with a combination of a time-distributed 2D Convolutional Neural Network (CNN) layers and an LSTM. The proposed setup boasts an impressive 92.65% accuracy on a demanding dataset due to poor video quality and multiple people present in the videos.

### 2.2 Identified XAI Models

Through a review of the literature, the following XAI models were identified for potential application in pose-based cricket shot classification. These models differ in their explanation type, whether they explain individual predictions or overall model behavior, and in their compatibility with different model architectures. A brief overview of the relevant categories is provided below:

- **Local vs. Global Explanations**
  - *Local explanations* interpret the model's decision for a specific instance (e.g., why was this **particular** cricket shot classified as a pull shot?).
  - *Global explanations* provide an overall understanding of the model's behavior across the entire dataset (e.g., which features are most important for shot classification **overall**?).
- **Model-Agnostic vs. Model-Specific**
  - *Model-agnostic* methods can be applied to **any machine learning model**. These treat the model as a black box and use perturbations or surrogate models for interpretation.

- *Model-specific* methods are tailored to a **specific class of models** (e.g., CNNs or decision trees) and leverage internal model components such as gradients or tree paths.

The selected XAI methods, categorized based on these dimensions, are described below:

- **Feature Importance** – A global, model-specific method primarily used in tree-based models like Random Forests. It is computed during training by measuring how much each feature contributes to reducing impurity (e.g., Gini index or entropy) at decision tree splits. This means features that are used often and split data effectively receive higher importance scores. However, it can be biased toward features with many categories or higher cardinality and does not capture the true impact on model performance or offer local explanations.
- **Permutation Importance** – Introduced by Altmann et al. [2010], it is a global, model-agnostic method calculated after model training. It works by randomly shuffling/permuting the values of a single feature and measuring how much the model’s performance degrades based on metrics (e.g. F1 score, accuracy). A significant drop in accuracy indicates high feature importance. This method gives a more direct measure of a feature’s influence on the model’s predictions but can be affected by correlated features and requires more computation. Unlike feature importance, it evaluates how much the model truly relies on a feature rather than just how frequently it is used.
- **SHAP (SHapley Additive Explanations)** – A game-theoretic, model-agnostic approach originally introduced by Lundberg and Lee [2017], Lundberg et al. [2020], SHAP is one of the most effective and commonly used XAI models thanks to its high versatility. It works on a similar principle to permutation importance, where by altering values of each feature and observing its impact on the final result, it computes its average marginal contribution across all feature subsets. SHAP supports both local and global explanations and is highly efficient for tree-based models via TreeSHAP.
- **Grouped SHAP** – An extension of SHAP that aggregates attributions within semantically related feature groups (e.g., joint coordinates of the same body part like RightElbow\_x, RightElbow\_y, RightElbow\_z presented as RightElbow). This approach enhances interpretability in high-dimensional datasets, as demonstrated in XPose Qiu et al. [2024].
- **LIME (Local Interpretable Model-agnostic Explanations)** – Introduced by Ribeiro et al. [2016], LIME is a local, model-agnostic technique that perturbs the input around a given instance and fits an interpretable surrogate model (e.g., linear model) to approximate the black-box model locally.
- **ALE (Accumulated Local Effects) Plots** – A model-agnostic, global explanation method introduced by Apley and Zhu [2020] and further popularized in interpretable ML literature Molnar [2025]. ALE plots estimate the average effect of a feature on the model’s prediction while accounting for the actual distribution of the data. Unlike Partial Dependence Plots (PDP), which can suffer from unrealistic extrapolations in areas with little or no data, ALE plots restrict their analysis to regions where data is present, making them more robust and interpretable when features are correlated. The method works by dividing the feature space into intervals and calculating the local effect of a feature within each bin, then accumulating these effects across the feature’s range. This allows for a more faithful representation of a feature’s influence without assuming independence from other features. ALE plots are especially useful when one wants to understand the direction and magnitude of a feature’s impact across its value range, though they do not provide local explanations.
- **Grad-CAM (Gradient-weighted Class Activation Mapping)** – A model-specific, local explanation technique designed for convolutional neural networks (CNNs), originally introduced by Selvaraju et al. [2020]. Grad-CAM provides visual explanations by generating class-specific heatmaps that highlight the spatial regions in the input (typically images or frames) that were most influential in the model’s prediction. It works by computing the gradients of a target class score with respect to the activations of the last convolutional layer. These gradients are then globally averaged to obtain importance weights for each feature map, which are combined with the feature maps to produce a coarse localization map. Grad-CAM is particularly valuable in applications like image classification and human pose estimation, where interpretability must be spatially grounded. While it offers intuitive visual insights and is easy to understand, its resolution is limited by the size of the convolutional

Table 1: XAI Model Categorization

Model	Explanation Type	Compatibility
SHAP	Local + Global	Model-agnostic
Grouped SHAP	Local + Global	Model-agnostic
LIME	Local	Model-agnostic
Permutation Importance	Global	Model-agnostic
ALE Plots	Global	Model-agnostic
Grad-CAM	Local	Model-specific (CNNs)
RISE	Local	Model-agnostic
Anchors	Local	Model-agnostic

feature maps and it only supports CNN-based architectures. Despite this, it remains one of the most widely used tools for debugging CNN predictions, especially in fields like medical imaging, sports analytics, and autonomous driving.

- **RISE (Randomized Input Sampling for Explanation)** –Introduced by Petsiuk et al. [2018], RISE is a model-agnostic technique that explains image-based predictions by figuring out which parts of the input were most important for the model’s decision. It does this by covering different areas of the image in random patterns and observing how the prediction changes. If hiding a certain region causes the prediction confidence to drop, that area is considered important. By repeating this many times with different patterns, RISE creates a visual heatmap that shows which parts of the image influenced the decision the most.
- **Anchors** – A model-agnostic, local explanation technique that explains individual predictions using simple IF-THEN rules. These rules, called “anchors,” describe the conditions under which the model makes the same prediction with high confidence. For example, an anchor for classifying an image as a “pull shot” might be: IF the left elbow is at location x1 AND the head is at location x2, THEN the model will predict “pull shot” with high precision. The idea is to find the minimum set of conditions that, when true, almost always lead to the same prediction. Anchors are easy to understand and can be useful for users who want clear, rule-based insights into why the model made a certain decision Ribeiro et al. [2018].

Table 1 gives a summary of the aforementioned models.

These methods span a variety of explanation strategies, including feature attribution, visual saliency, perturbation analysis, and rule extraction. Evaluating XAI methods is difficult because they cannot be evaluated quantitatively effectively, as they do not provide results but explain them. Hence, in this paper, the XAI methods will be evaluated qualitatively using Interpretability, Usefulness, Visualization Clarity, Domain Relevance and Computational Simplicity.

### 3 Experimental analysis of applying different XAI techniques

To evaluate and compare different XAI methods for cricket shot classification, two parallel pipelines were constructed: one using pose keypoint data with a Random Forest classifier, and another using a CNN-LSTM model applied to raw video frames. Each pipeline was chosen to align with the requirements of the XAI methods being tested. Figure 1 shows a diagram of the implementation pipeline.

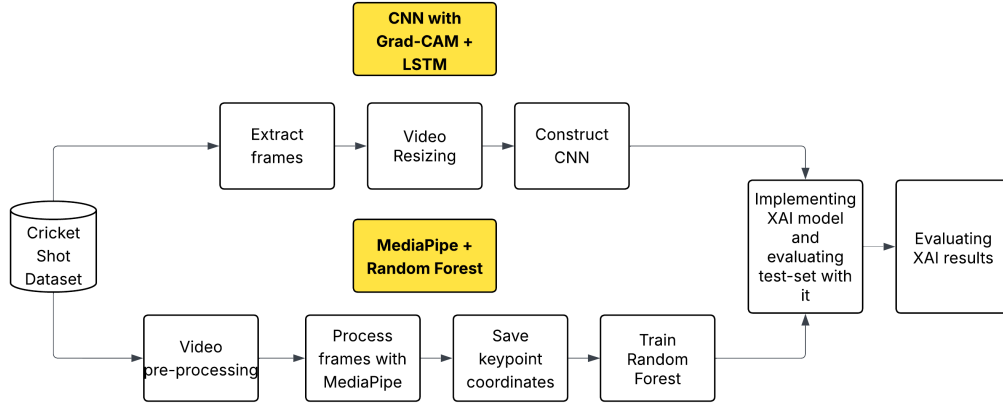


Figure 1: Diagram showing the implementation pipeline for this study. Each stage—from the unprocessed video dataset to the final XAI explanations—is annotated with the corresponding method used.

### 3.1 MediaPipe-Random Forest + Feature Importance, Permutation Importance, SHAP, LIME and ALE

This pipeline involves extracting pose keypoints from cricket shot videos using MediaPipe, followed by classification using Random Forest and interpretation through various XAI techniques.

1. A dataset of 722 labeled cricket shot videos is used.
2. Pre-processing is performed on the videos to ensure MediaPipe extracts the keypoints of the correct person. More information about this can be found in subsection 5.1.
3. Each video is processed frame-by-frame using MediaPipe, which extracts the  $x$ ,  $y$ ,  $z$  coordinates of 17 upper-body keypoints, as referenced in Siddiqui et al. [2023].
4. Two variations of keypoint data are stored:
  - A single-row-per-video CSV, where keypoint values are averaged across all frames.
  - A full-frame CSV, where each frame with detected keypoints is stored as a separate row (resulting in 17,000 entries).
5. A Random Forest classifier is trained in two ways:
  - Multi-class classification with 5 classes (Bowled, Cover Drive, Defence, Pull Shot, Reverse Sweep).
  - Binary classification per shot type (e.g., Bowled vs. Not Bowled).
6. All models are trained with an 80/20 train-test split, using `n_estimators=100` and `random_state=42`.
7. Once trained, the following XAI methods are applied:
  - **Feature Importance** and **Permutation Importance** via scikit-learn.
  - **SHAP**, **LIME**, and **ALE plots**, which are compatible with tree-based models and readily implemented.

### 3.2 CNN-LSTM + Grad-CAM Pipeline

This pipeline operates directly on raw video frames and utilizes a CNN-LSTM architecture based on Ahmad et al. [2023], allowing for temporal and spatial modeling. Grad-CAM is applied to visualize important spatial regions in the input.

1. The dataset of cricket shot videos is used, and each video is split into 10 evenly spaced frames. Each frame is resized to  $100 \times 100$  pixels and retains all 3 RGB channels.
2. Videos are labeled according to their corresponding shot type.

3. The CNN-LSTM model architecture is described in Table 2.
4. After training, Grad-CAM is applied to the final convolutional layer (`target_layer`) to generate heatmaps showing which regions of the input frames contributed most to the model's decision.
5. These Grad-CAM heatmaps are overlaid on the frames to visually inspect and interpret the model's reasoning.

All the experiments were carried out in Jupyter Notebooks. All the computations were performed on my personal computer equipped with an AMD Radeon RX 9070 XT GPU, Ryzen 9700X CPU, and 64GB of RAM.

## 4 Results

After both the CNN and the RF had been trained, XAI methods were implemented on top of them. The selected XAI methods, namely Feature Importance, Permutation Importance, SHAP, Grouped SHAP, LIME, Grad-CAM, and ALE plots, were implemented. Anchors were not implemented due to time constraints and the overabundance of features, and RISE was not implemented due to its high similarity to Grad-CAM. This section will present a comparative analysis of the outcomes from the different XAI techniques applied to the cricket shot classification task, as well as pinpointing interdependencies. This will be done by evaluating and comparing the results of XAI explanations. All XAI results presented, except for those from Grad-CAM, are derived from binary classification models. This choice stems from the superior performance of the Random Forest classifier when trained in a one-vs-rest (binary) setting compared to a multi-class setup. Additionally, global explanation methods (e.g., Feature Importance, Permutation Importance, SHAP) applied to the multi-class model yielded less informative insights for addressing the research sub-questions. While multi-class explanations do highlight generally important features across all classes, the binary classifiers offered more focused and actionable interpretations for individual shot types, making them more appropriate for this study's analytical objectives.

### 4.1 XAI model results analysis

For each XAI model, an evaluation of its Interpretability, Usefulness, Ability to pinpoint interdependencies, and Visual Clarity of explanations (In what ways can the information be displayed, e.g., plots, text) will be performed, as well as some implementation notes focusing on its ease of use and libraries.

#### 4.1.1 Feature Importance

Implemented on RF by evaluation through training. Example in Figure 6

**Interpretability:** Provides feature importances in plain text highlighting which features were overall most influential in classifying a shot under a certain type. Easily understandable for humans with a general overview of the result meanings.

**Usefulness:** Both the binary-classification and multi-class classification results can be used by players/coaches/analysts. In the binary classification, the most impactful features for each shot type give the viewer a good guide on which body keypoints to focus on (e.g., the `RightElbow_y` has high importance, meaning the focus should be on keeping the elbow's vertical position constant during the shot). It is not useful for analyzing a specific shot because it only has global explanations.

**Interdependencies:** Due to being implemented directly in the scikit-learn library, it cannot be altered to observe interdependencies. This can be done, however, by altering the input data and seeing the effect on the overall explanation. This is potentially very time-consuming.

**Visual Clarity:** Resulting values are in plain-text, however, if converted into a bar-chart as can be seen in Figure 6, results become more intuitive.

**Implementation Notes:** Straightforward to implement if scikit's Random Forest implementation is used by simply calling a method of the RF.



#### 4.1.2 Permutation Importance

Implemented on RF post-training on test data. Example in Figure 7

**Interpretability:** Provides feature importances in plain text in the same manner as Feature Importance.

**Usefulness:** Again, it has the same strengths and weaknesses as Feature Importance; both can be used to cross-check resulting feature importances.

**Interdependencies:** Can be used for finding interdependencies through altering the test data and seeing the effect on the overall explanation, same as Feature Importance.

**Visual Clarity:** The resulting values provide plain-text results easily displayable and interpretable for anyone with general knowledge of its workings.

**Implementation Notes:** Straightforward to implement if scikit's classifier models are used, due to scikit having an implementation of Permutation Importance. Computationally efficient.

#### 4.1.3 SHAP

Implemented on RF for both multi-class and binary classification settings. Example in Figure 8

**Interpretability:** SHAP provides both global and local feature attributions, making it one of the most interpretable XAI methods. It assigns a numerical weight to each feature, indicating its contribution to the final prediction. The greater the weight, the greater the importance, as in Feature Importance.

**Usefulness:** Very useful in cricket shot classification. SHAP explains individual predictions and also highlights which features consistently matter across many predictions. This helps identify critical joints (e.g., right shoulder, left elbow) for different shot types, and for analyzing specific instances of shots.

**Interdependencies:** SHAP's basic implementation does not capture feature interdependencies directly, but its local nature can help infer some dependencies by examining co-occurring important features.

**Visual Clarity:** SHAP plots are among the most intuitive and easy to understand: force plots and waterfall plots clearly show directionality of influence of features; summary plots aggregate insights across the dataset.

**Implementation Notes:** Simple to implement using the SHAP Python library. Compatible with scikit-learn RF models using TreeExplainer.

#### 4.1.4 Grouped SHAP (by Joint)

Implemented on RF using grouped feature attribution. Example in Figure 9

**Interpretability:** Groups the x, y, z coordinates of each joint into a single feature, making it easier to understand which body joint matters, instead of interpreting each coordinate separately.

**Usefulness:** Helps analysts focus on joint-level insights. For example, seeing that the "Right Elbow" overall has high importance could be more important than looking at separate x/y/z dimensions.

**Interdependencies:** While individual dependencies aren't shown, grouping helps reveal which joints matter in combination. Patterns across joints become easier to interpret.

**Visual Clarity:** Bar plots of grouped joint contributions make the results very digestible. Especially useful for communicating with non-technical stakeholders.

**Implementation Notes:** Requires manual preprocessing to group SHAP values. Supported via SHAP. Explainer with group definitions.

#### 4.1.5 Grouped SHAP (by Limb)

Implemented by logically grouping joints into limbs (e.g., left arm, right arm, head). Example in Figure 10

**Interpretability:** Very intuitive – coaches and players can directly relate explanations to body parts. Easier than joint-level grouping for high-level analysis.

**Usefulness:** Makes it possible to give limb-specific feedback. For example, the model may indicate the "left arm" was critical in the Reverse Sweep. Not very useful since only 3 groupings exist (head, left arm, right arm). If full-body MediaPipe were implemented, it could be more useful.

**Interdependencies:** Shows which limbs contribute together to prediction, but does not capture intra-limb dynamics.

**Visual Clarity:** Results in numerical values, can be converted to compact bar charts with few variables. Suitable for presentation-level insights.

**Implementation Notes:** Grouping requires manual definition of body parts and summing of joint-level SHAP values, similar to grouping by joint.

#### 4.1.6 LIME

Implemented on RF with individual test samples. Example in Figure 12

**Interpretability:** Provides local explanations by showing which features influenced a single prediction. The output is human-readable, not intuitive at first, and needs supplementary explanations.

**Usefulness:** Can be useful for exploring what features drove a single prediction. Weight values are generally very low (0.01 or 0.00), making them less useful than SHAP's local explanations.

**Interdependencies:** Can reveal patterns if analyzed across multiple samples, but no built-in way to analyze feature interactions.

**Visual Clarity:** Outputs plain-text of feature weights, only one option for visualization available other than manual visualizations.

**Implementation Notes:** Easy to implement with lime library in python. Non-deterministic; repeat runs give different results unless random seeds are fixed.

#### 4.1.7 ALE Plots

Implemented on RF for global explanations. Example in Figure 5

**Interpretability:** Plots how predictions change when a single feature changes, while averaging out the effects of other features. Easy to interpret with moderate explanation.

**Usefulness:** Can indicate whether a feature has a linear or nonlinear effect on predictions. Very time-consuming if an in-depth analysis of the feature needs to be done. Due to the high dimensionality of the input data, observations become sparse, meaning fewer data points exist in the neighborhood needed to calculate ALE values reliably.

**Interdependencies:** Can model feature interdependencies, but only the second-order effect, meaning the main effects of the features have been accounted for, so the interdependencies presented are close to zero. Also extremely time-consuming with a high-dimensional feature set.

**Visual Clarity:** Can only present 1-D or 2-D plots, meaning that when there is a high feature space, the features need to be inspected one by one. Outliers in the data can lead to uninterpretable plots.

**Implementation Notes:** Implemented via pyALE. Less useful with sparse or noisy data.

#### 4.1.8 Grad-CAM

Implemented on the CNN-LSTM model using the last convolutional layer. Example in Figure 4

**Interpretability:** Visual heatmaps highlight regions in the frame that contributed most to the prediction. Particularly intuitive for video/image-based models.

**Usefulness:** Very helpful in assessing whether the model is focusing on the batter or irrelevant background. Great for validating model behavior in spatial terms.

**Interdependencies:** Does not provide explicit joint/keypoint interdependencies, but can show whether multiple regions (e.g., head + bat) are attended simultaneously if properly trained

**Visual Clarity:** Has excellent clarity because overlays make it clear which part of the image was "important". Resolution depends on the convolutional layer size.

**Implementation Notes:** Requires accessing model internals and gradients. Needs careful identification of target layers.

In summary, SHAP and Grouped SHAP offered the most interpretable and context-relevant insights, particularly for understanding which keypoints influenced each shot, thanks to SHAP's ability to provide local and global explanations. Grad-CAM provided visual cues but lacked granularity at the keypoint level and proper training of the CNN, more information about this in subsection 5.2. While LIME was promising for exploring variation, it's inferior to SHAP thanks to its more intuitive visualizations and more significant feature weights. Permutation-based methods were helpful in global trends but did not capture localized interactions or feature dependencies as effectively. ALE plots were unstable and lacked highly interpretable results due to the high dimensionality of the input and regular anomalies.

## 4.2 Consistency Across Feature Attribution Methods

When comparing global explanations from Feature Importance, Permutation Importance, and SHAP (using TreeExplainer), we observe a strong consistency in the top-ranked features across all pipelines. Specifically:

- The top 4–6 features tend to be the same across methods, although their exact importance scores differ, as can be seen in Figure 6-8.
- The ranking diverges more in the bottom half of the feature set, likely due to noise, low variance, or weak correlation with shot type.

## 4.3 Identifying most influential keypoints

Using the resulting Global and Local explanations of Feature and Permutation Importance, SHAP, Grouped-SHAP, and Lime, conclusions can be made for the most influential keypoints in each shot type. An example of the analysis of Cover drive results is provided below. For the results of the other shots, please refer to the GitHub repository<sup>1</sup>.

As demonstrated in the global explanation methods, Feature Importance, Permutation Importance, and SHAP summary plots, LeftPinky\_y consistently emerges as the most influential keypoint, followed closely by RightPinky\_y, RightWrist\_y, and LeftWrist\_y as can be seen in Figure 6-8. This ranking aligns with the expected biomechanical relevance of these joints, as both hands hold the bat during a cover drive shot, making their vertical trajectories critical in the model's decision-making.

The cover drive is a cricket shot characterized by a fluid, arcing follow-through. While the bat travels forward and slightly upward, the batter's body lowers during execution to maintain balance and positioning. The wrists and pinkies retain a relatively consistent vertical position throughout the stroke. Such stability is captured in the SHAP analysis: LeftPinky\_y values that are lower tend to strongly support the classification of a cover drive, whereas higher values negatively impact the prediction. In contrast, RightPinky\_y exhibits a more nuanced pattern. Mid-to-low values contribute positively, but high values yield mixed effects, either supporting or slightly detracting from classification accuracy.

These insights can then be confirmed through local explanations, including LIME and Waterfall SHAP plots as can be seen in Figure 11-14, reinforcing the critical role of bat-handling keypoints in distinguishing the cover drive from other shot types.

## 4.4 Keypoint interdependencies

When extended with global explanations and Grouped SHAP (by joint and by limb), SHAP's local explanations provide a deeper view of feature relationships. Key findings include:

- In binary classifiers, in all shot types, features of high importance always include the right and left pinky vertical positions, pointing out a possible interdependence between them.

---

<sup>1</sup><https://github.com/bruno-martinovic/XAI-in-cricket-shot-classification.git>

This makes sense as both pinkies are the body’s keypoints directly connected to the bat and most closely follow the bat’s movement.

- Observing the local SHAP explanations of high prediction probability and high single feature importance, interdependencies between naturally connected body parts (e.g., wrist and pinky, elbow and wrist) can be observed. This is most visible in wrist and pinky SHAP values, where if one of them is highly influential, there is a high likelihood that the other is as well, as can be seen in Figure 13
- Grouping by limb (e.g., "Right Arm") highlights that shot classification often depends on the relative importance of one limb over another. For example, in all shot types except for Reverse Sweep, the right arm is more important in classification; however, in Reverse Sweep, both the left and right arm are of almost equivalent importance, the left even edging out the right slightly as can be seen in Figure 2. This again is a logical result since in the reverse sweep, the shot is taken from an opposite stance.

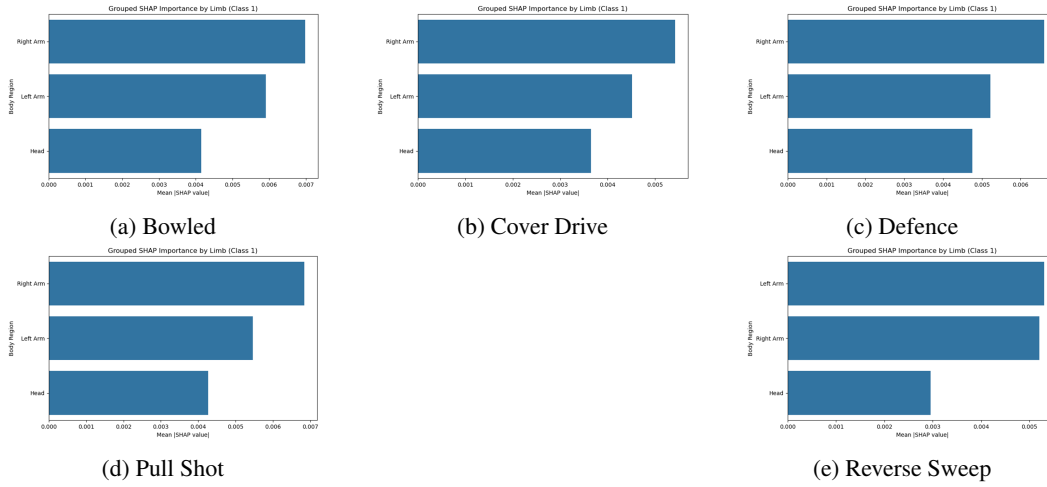


Figure 2: Grouped SHAP importance by body region across five cricket shot classes.

Overall, keypoint interdependencies can be observed in both local and global explanations, but local explanations provide a much more insightful image. While this requires inspecting many local explanations to find patterns, a lot of them can be found by inspecting the samples with the highest prediction probability and most influential SHAP values. For this purpose, SHAP performs best. A combination of Feature/Permutation importance + LIME can be used; however, LIME’s feature weights are less precise than SHAP’s and generally carry less weight, so they are therefore harder to discern patterns from.

#### 4.5 Using SHAP’s results

This paper compares the usefulness of different XAI methods from a technical standpoint and does not delve deeply into presenting the explanations to people outside of the computer science domain. For a detailed overview of how SHAP explanations can be used for this purpose, I would recommend reading the paper Kumar [2025], by Ansh, a member of my research project group. He delves into explaining cricket shots across three different expertise levels with the help of SHAP values.

## 5 Limitations

This section outlines the key challenges and limitations encountered during the course of the project. A range of issues, varying in scope and severity, impacted the number and depth of possible implementations and modifications within the available timeframe. The most significant and time-consuming challenges are summarized below.

## 5.1 Problems with MediaPipe

To feed the videos into MediaPipe, they had to be preprocessed, because, when used without processing, MediaPipe misinterpreted the images and took the wrong person's keypoints, as can be seen in Figure 3a, an example of correct interpretation can be seen in Figure 3b. In order to ensure MediaPipe interpreted the correct person, the videos were trimmed by 0.5 seconds at the start and 1 second at the end, since in these parts of the video, the bowler and the flight of the ball are shown instead of the batter. Then, Ultralytics YOLOv8's object detection was used to feed a cropped version of the image to MediaPipe by identifying the batter and then cropping the frame to show only him in the image. This resulted in MediaPipe providing correct coordinates for keypoints, as can be seen in Figure 3c.

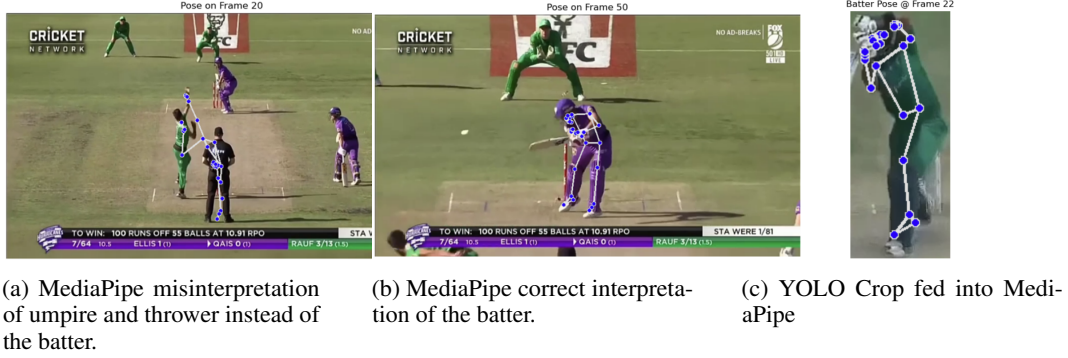


Figure 3: Comparison of MediaPipe's pose estimation results: incorrect vs. correct subject detection.

## 5.2 Grad-CAM results

While Grad-CAM gave significant insight into the regions of the image the CNN focuses on, these results turned out to be mixed. From a classification standpoint, the CNN + LSTM classifier performed great with a 0.9241 validation accuracy. However, while the results were very accurate, using Grad-CAM made it clear that the CNN focused on wrong regions of the image as can be seen in Figure 4. This is likely to a lack of CNN training on human poses. The CNN focuses on elements such as the scoreboard and grass type on the side of the hitter since most shots of the same type were taken on the same fields or in the same games. The implementation of Grad-CAM on this CNN, hence, does not give any insight into interdependencies of keypoints or which keypoints are essential to which shot type. Multiple attempts were made at implementing Grad-CAM on pose estimation models. However, most of the models could not be implemented due to being open-source and using deprecated features, and I failed to modify the ones that could be implemented so that they could support Grad-CAM due to their high level of complexity and time constraints.

## 6 Responsible Research

### 6.1 Reproducibility and Transparency

To ensure the reproducibility of this experiment, all the code used for the final results has been documented and split into logical groups and cells. Documentation for running and understanding

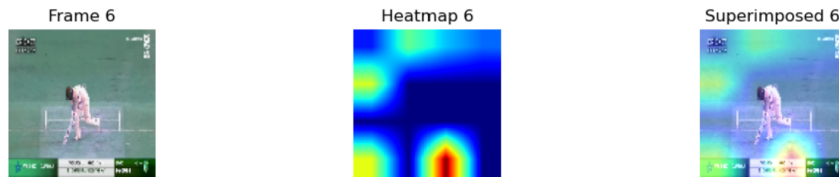


Figure 4: Grad-CAM result on Frame 6 of the video clip

the code has been done through an introductory README file and comments throughout the code itself in the notebooks, as well as titles and subtitles inside the notebooks. All code has been made publicly available on GitHub<sup>2</sup>.

The pose estimation, classification, and explanation models were implemented using popular open-source libraries (MediaPipe, OpenCV, scikit-learn, SHAP, PyALE, PyTorch, etc.) to ensure accessibility. Fixed seeds were used where applicable so that the results could be reproduced.

## 6.2 Ethical Considerations

This study did not involve human subjects, and therefore, no ethical approval was required. All data sources were publicly available and used in accordance with their terms of use; the source of the video dataset has been referenced as requested.

This paper contributes to making AI analysis in cricket more accessible and interpretable. It presents how XAI can be applied to cricket shot classification and provides a template that can be used in the analysis of other sports.

The conclusions drawn in this paper could turn out to be incorrect due to the wrong implementation of code or misinterpretation of the results. This can result in readers getting the wrong image about the usefulness of XAI in cricket shot classification. To mitigate this problem, the logic behind all conclusions has been explained, and all the figures and code are accessible, so that readers can draw their own conclusions.

## 7 Discussion, Conclusion, and Future Work

Through this research, I expanded the use of XAI in analyzing cricket shot techniques by combining pose-based cricket shot classification with interpretability techniques from XAI. This led to the formulation of potentially viable ways of using XAI for analysis, coaching, or training purposes.

### 7.1 XAI Model recommendations

From this research, I concluded that SHAP is the best XAI model among the tested for cricket shot analysis overall due to its flexibility in explanation visualizations and observation of feature interdependencies through its local and global explanations. The effectiveness of SHAP observed in this study mirrors findings in other domains. For instance, SHAP has also been successfully used in posture classification in medical contexts Dindorf et al. [2021] and performance analysis in basketball Ou-Yang et al. [2025]. These parallels suggest that the selection of SHAP as the primary method is not only domain-relevant but also broadly supported by XAI applications in related fields. SHAP is also model-agnostic, making it suited for all AI models, so if a better classification model is found than the ones used in this paper, SHAP can be used on it.

A possible XAI model that can be used alongside SHAP is Grad-CAM. While the results in this paper were not insightful in human body keypoint analysis, if Grad-CAM were to be applied to a properly trained CNN for human-pose estimation, its heatmap overlayed over the image of the shot could be insightful, especially if combined with SHAP.

### 7.2 Highlighting keypoint contributions and interdependencies

The second research subquestion examined the feasibility of identifying individual keypoint contributions and the interdependencies between them in the context of cricket shot classification. By leveraging SHAP's local explanations validated through comparison with LIME, and SHAP's global explanation validated through comparison with Feature Importance and Permutation Importance, it was possible to identify the most influential keypoints for each shot type and to reveal meaningful interdependencies. Across all shot categories, the most significant keypoints were consistently found in the lower arm region, specifically between the elbow and the hand. Interdependencies were primarily observed between anatomically adjacent keypoints; however, the observed dependency between left and right hand keypoints was a notable exception. While not anatomically connected, these keypoints are both in contact with the bat, thus exhibiting functional linkage during shot execution.

---

<sup>2</sup><https://github.com/bruno-martinovic/XAI-in-cricket-shot-classification.git>

### 7.3 Future Work

Expanding upon the current implementation and evaluation of XAI methods in cricket shot classification, several approaches can be taken to broaden the scope, applicability, and robustness of this research:

#### 7.3.1 Dataset Expansion

The dataset used in the paper includes five distinct shot types and a small number of videos (722). Future work should focus on expanding the dataset in multiple ways:

1. A broader variety of shot types (e.g., straight drive, hook, cut, helicopter) to increase classification complexity and applicability.
2. Processing based on dominant hand and skill level to improve the model's generalizability and accuracy.
3. A greater volume of videos and videos of higher quality and more focused on the batter. As observed in the Limitations section, MediaPipe had problems with correctly estimating the pose of the batter due to the fact that there were too many other participants in the shot and the low resolution of the videos.

This would improve the predictive performance of classification and test the scalability and stability of XAI explanations across a wider range of inputs.

#### 7.3.2 Alternative classification models and multi-class focus

In this paper, I focused more on binary classification results because they performed better, and XAI provided more useful explanations when applied to them. Therefore, research could be done on other classification models, which could perform better or similarly in multi-class to binary classification. This is closely connected to the following recommendation for future works ("Integration of other XAI models").

#### 7.3.3 Integration of other XAI models

While this paper covered some of the most famous and commonly used XAI models, future research could benefit from incorporating other XAI models, such as RISE, Anchors, and Temporal saliency methods, to see how their performance compares. Exploring these can complement the current findings and potentially offer more in-depth or reliable explanations.

#### 7.3.4 Human-Centered Evaluation of XAI Outputs

While Kumar [2025] explores the usability of XAI outputs from a user-centered perspective, primarily focusing on how coaches and athletes interpret and react to explanations, future work could extend this line of inquiry by adopting a more technical lens. Specifically, research could investigate:

1. Systematic benchmarking of explanation quality, using objective metrics such as fidelity, sparsity, and stability of the explanations.
2. Expert evaluation of explanations to assess their alignment with textbook definitions of cricket shot techniques.

By complementing human usability studies with rigorous technical evaluation, future research can better assess the practical and scientific utility of XAI in pose-based sports analytics.

## References

- Waqas Ahmad, Muhammad Munsif, Habib Ullah, Mohib Ullah, Alhanouf Abdulrahman Alsuwailam, Abdul Khader Jilani Saudagar, Khan Muhammad, and Muhammad Sajjad. Optimized deep learning-based cricket activity focused network and medium scale benchmark. *Alexandria Engineering Journal*, 73:771–779, 2023. ISSN 1110-0168. doi: <https://doi.org/10.1016/j.aej.2023.04.062>. URL <https://www.sciencedirect.com/science/article/pii/S1110016823003368>.
- André Altmann, Laura Tološi, Oliver Sander, and Thomas Lengauer. Permutation importance: a corrected feature importance measure. *Bioinformatics*, 26(10):1340–1347, 04 2010. ISSN 1367-4803. doi: 10.1093/bioinformatics/btq134. URL <https://doi.org/10.1093/bioinformatics/btq134>.
- Daniel W. Apley and Jingyu Zhu. Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(4): 1059–1086, 2020. doi: <https://doi.org/10.1111/rssb.12377>. URL <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssb.12377>.
- Ks Sreekar Datta, G Narasimha Naidu, Rahul Dasari, and T V Jayakumar. Advancements in cricket shot detection: Integrating human pose estimation and deep learning for automated analysis. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages 1–7, 2024. doi: 10.1109/ICCCNT61001.2024.10726249.
- Carlo Dindorf, Jürgen Konradi, Claudia Wolf, Bertram Taetz, Gabriele Bleser, Janine Huthwelker, Friederike Werthmann, Eva Bartaguiz, Johanna Kniepert, Philipp Drees, Ulrich Betz, and Michael Fröhlich. Classification and automated interpretation of spinal posture data using a pathology-independent classifier and explainable artificial intelligence (xai). *Sensors*, 21(18), 2021. ISSN 1424-8220. doi: 10.3390/s21186323. URL <https://www.mdpi.com/1424-8220/21/18/6323>.
- Ansh Sharma Kumar. Generating expertise specific explanations in cricket pose estimation: design, implementation, and evaluation of adaptive xai feedback. Bachelor’s thesis, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, June 2025. URL <http://repository.tudelft.nl/>. CSE3000 Research Project.
- Ashok Kumar, Javesh Garg, and Amitabha Mukerjee. Cricket activity detection. In *International Image Processing, Applications and Systems Conference*, pages 1–6, 2014. doi: 10.1109/IPAS.2014.7043264.
- Ryunosuke Kurose, Masaki Hayashi, Takeo Ishii, and Yoshimitsu Aoki. Player pose analysis in tennis video based on pose estimation. In *2018 International Workshop on Advanced Image Technology (IWAIT)*, pages 1–4, 2018. doi: 10.1109/IWAIT.2018.8369762.
- Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, page 4768–4777, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.
- Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. From local explanations to global understanding with explainable ai for trees. *Nature Machine Intelligence*, 2(1):56–67, Jan 2020. ISSN 2522-5839. doi: 10.1038/s42256-019-0138-9. URL <https://doi.org/10.1038/s42256-019-0138-9>.
- Josue Angel Mauricio Salazar and Hugo Alatrística-Salas. Football penalty kick prediction model based on kicker’s pose estimation. In *Proceedings of the 2024 9th International Conference on Machine Learning Technologies, ICMLT ’24*, page 196–203, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400716379. doi: 10.1145/3674029.3674061. URL <https://doi-org.tudelft.idm.oclc.org/10.1145/3674029.3674061>.
- Rahul Mili, Nayan Ranjan Das, Arjun Tandon, Saquelain Mokhtar, Imon Mukherjee, and Goutam Paul. Pose recognition in cricket using keypoints. In *2022 IEEE 9th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, pages 1–5, 2022. doi: 10.1109/UPCON56432.2022.9986481.



Christoph Molnar. *Interpretable Machine Learning*. 3 edition, 2025. ISBN 978-3-911578-03-5. URL <https://christophm.github.io/interpretable-ml-book>.

Yan Ou-Yang, Wei Hong, Liming Peng, Cheng-Xi Mao, Wen-Jia Zhou, Wei-Tao Zheng, Quan Wang, Feng Qi, Xue-Wei Li, Shi-Huan Chen, Ce Xu, and Yu-Fan Wang. Explaining basketball game performance with shap: insights from chinese basketball association. *Scientific Reports*, 15, 04 2025. doi: 10.1038/s41598-025-97817-3.

Vitali Petsiuk, Abir Das, and Kate Saenko. Rise: Randomized input sampling for explanation of black-box models. In *British Machine Vision Conference (BMVC)*, 2018.

Luyu Qiu, Jianing Li, Lei Wen, Chi Su, Fei Hao, Chen Jason Zhang, and Lei Chen. Xpose: explainable human pose estimation, 2024. URL <https://arxiv.org/abs/2403.12370>.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, page 1135–1144, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450342322. doi: 10.1145/2939672.2939778. URL <https://doi-org.tudelft.idm.oclc.org/10.1145/2939672.2939778>.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Anchors: high-precision model-agnostic explanations. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence, AAAI'18/IAAI'18/EAAI'18*. AAAI Press, 2018. ISBN 978-1-57735-800-8.

Ramprasaath R. Selvaraju, Michael Cogswell, Das Abhishek, Vedantam Ramakrishna, Parikh Devi, and Batra Dhruv. Grad-cam: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, 128(2):336–359, 02 2020. URL <https://www-proquest-com.tudelft.idm.oclc.org/scholarly-journals/grad-cam-visual-explanations-deep-networks-via/docview/2352397845/se-2>. Copyright - International Journal of Computer Vision is a copyright of Springer, (2019). All Rights Reserved; Last updated - 2024-11-21.

Hafeez Ur Rehman Siddiqui, Faizan Younas, Furqan Rustam, Emmanuel Soriano Flores, Julián Brito Ballester, Isabel de la Torre Diez, Sandra Dudley, and Imran Ashraf. Enhancing cricket performance analysis with human pose estimation and machine learning. *Sensors*, 23(15), 2023. ISSN 1424-8220. doi: 10.3390/s23156839. URL <https://www.mdpi.com/1424-8220/23/15/6839>.

## A CNN Setup

Table 2: Details of the CNN-LSTM model based on Ahmad et al. [2023]

Layer	Kernels	Size	Padding	Strides	Activation	Output Shape
TD (Conv2D) <sub>1</sub>	16	3x3	same	2x2	ReLU	10, 100, 100, 16
TD (MaxPooling2D) <sub>1</sub>	1	4x4	same	2x2	-	10, 25, 25, 16
TD (Conv2D) <sub>2</sub>	32	3x3	same	2x2	ReLU	10, 25, 25, 32
TD (MaxPooling2D) <sub>2</sub>	1	4x4	same	2x2	-	10, 6, 6, 32
TD (Conv2D) <sub>3</sub>	64	3x3	same	2x2	ReLU	10, 6, 6, 64
TD (MaxPooling2D) <sub>3</sub>	1	2x2	same	2x2	-	10, 3, 3, 64
TD (Conv2D) <sub>4</sub>	128	3x3	same	2x2	ReLU	10, 3, 3, 128
TD (MaxPooling2D) <sub>4</sub>	1	2x2	same	2x2	-	10, 1, 1, 128
TD (Flatten)	-	-	-	-	-	10, 128
LSTM	-	-	-	-	None	128
Dense	-	-	-	-	ReLU	64

## B XAI Results for Cover Drive

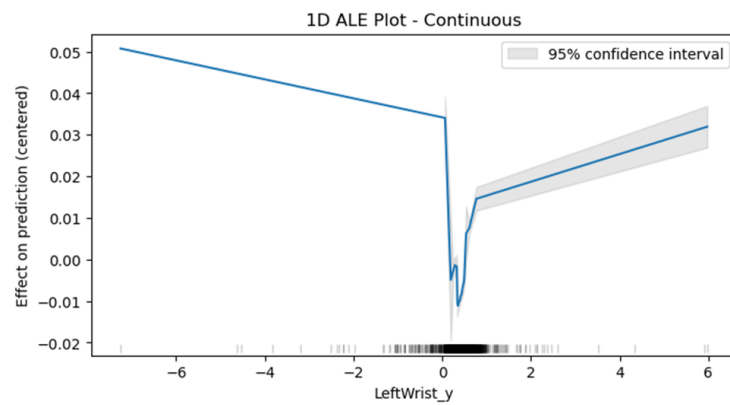


Figure 5: Accumulated Local Effects (ALE) plot showing the marginal effect of each keypoint feature for the cover drive class.

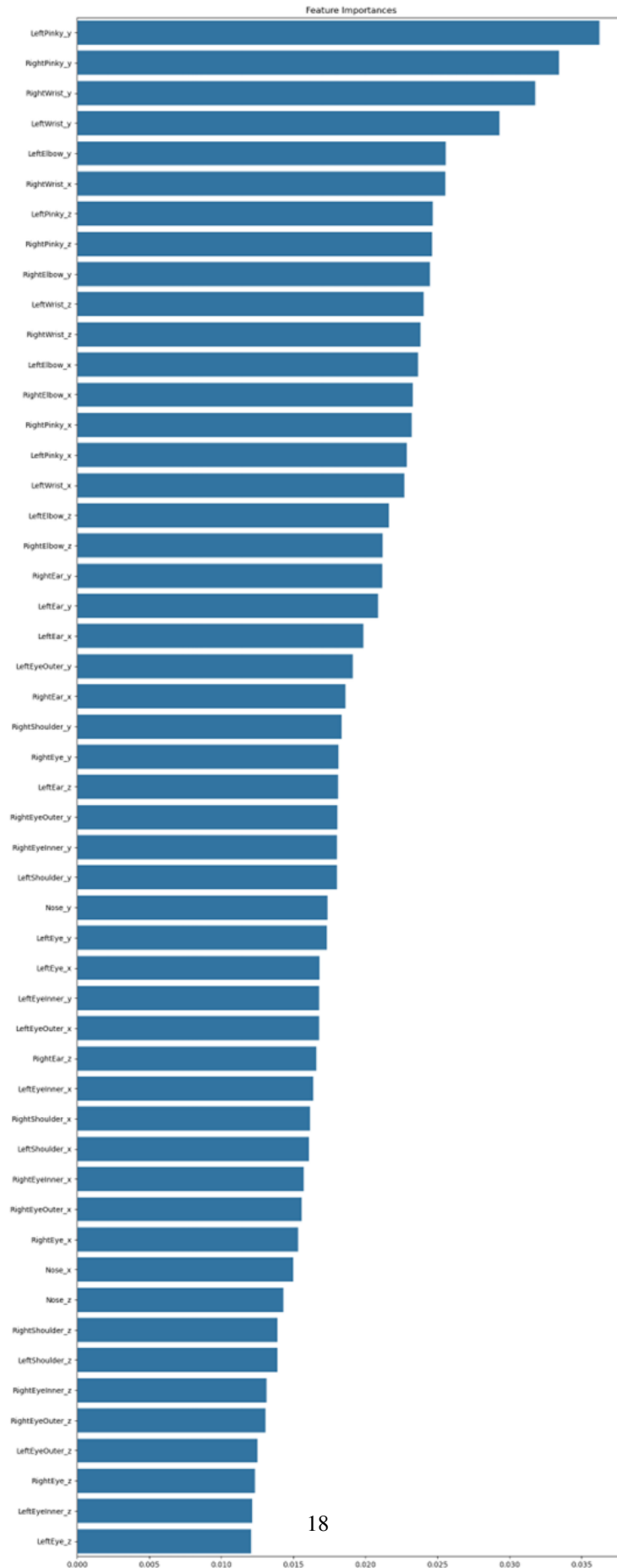


Figure 6: Feature Importance plot based on the Random Forest classifier for cover drive predictions.

```

LeftPinky_y: 0.0155
RightPinky_y: 0.0151
RightWrist_y: 0.0103
LeftWrist_y: 0.0062
RightElbow_y: 0.0041
RightElbow_x: 0.0040
LeftPinky_x: 0.0040
RightElbow_z: 0.0039
LeftElbow_y: 0.0038
RightPinky_x: 0.0037

```

Figure 7: Permutation Importance plot for the cover drive class, showing performance drop after feature shuffling.

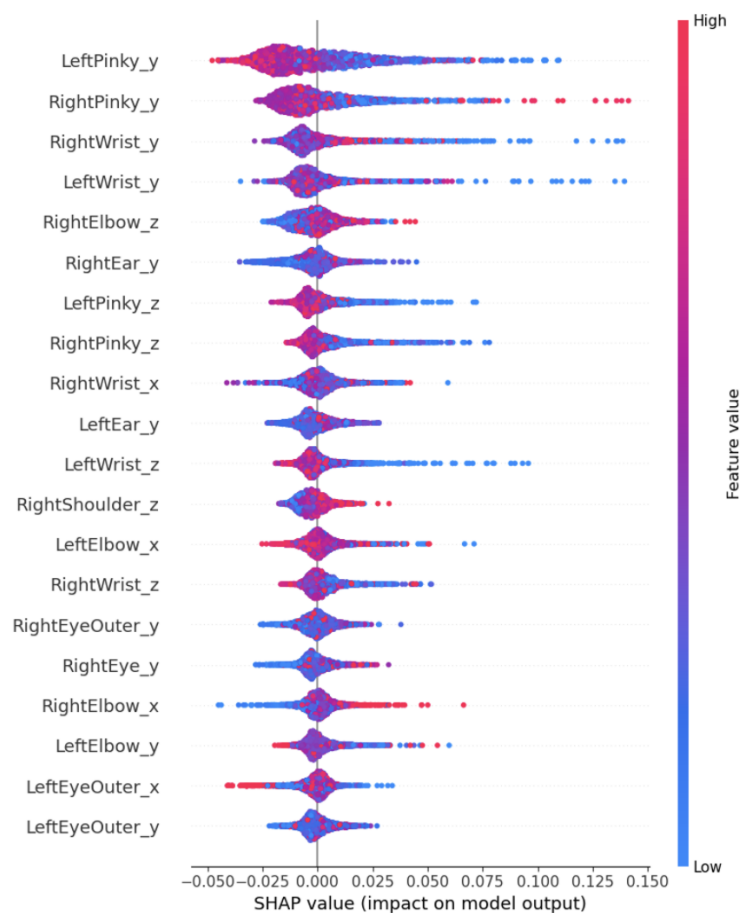


Figure 8: SHAP summary plot for the cover drive class showing average impact and value distribution of keypoints.

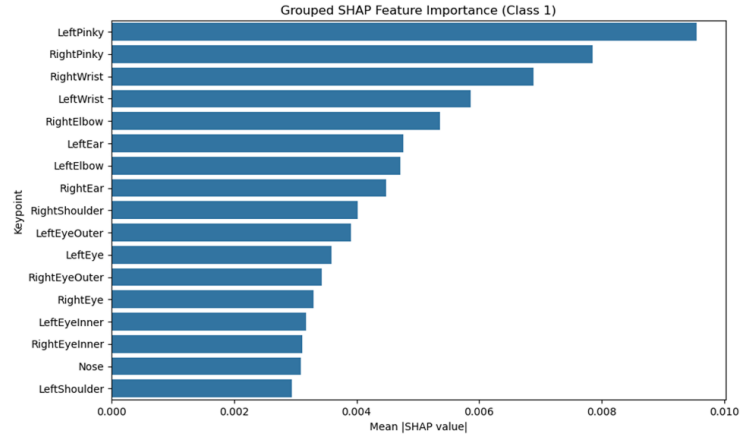


Figure 9: Grouped SHAP values by joint, showing which body joints had the highest average contribution to predicting a cover drive.

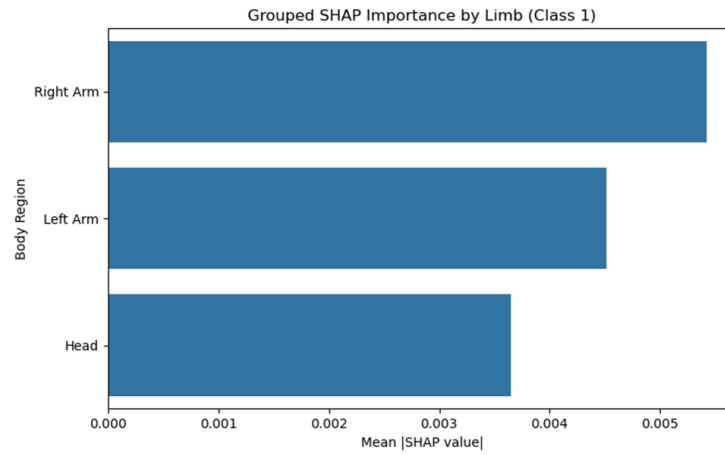


Figure 10: Grouped SHAP values by limb group, highlighting which regions of the body most influenced the cover drive classification.

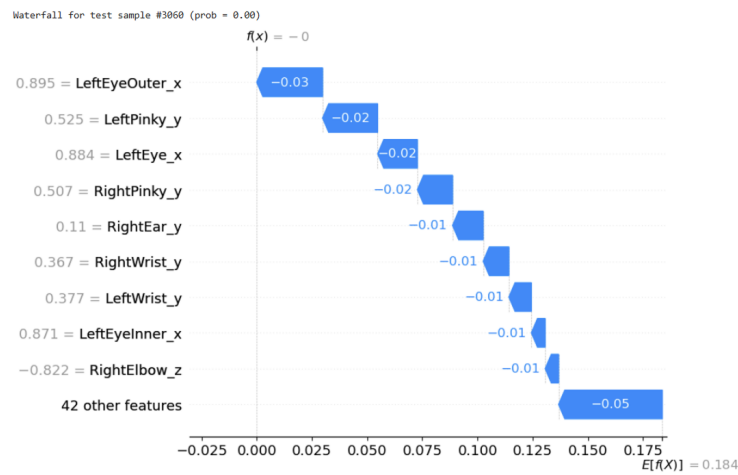


Figure 11: SHAP waterfall plot (1) showing keypoints with the most negative contribution to a cover drive prediction.

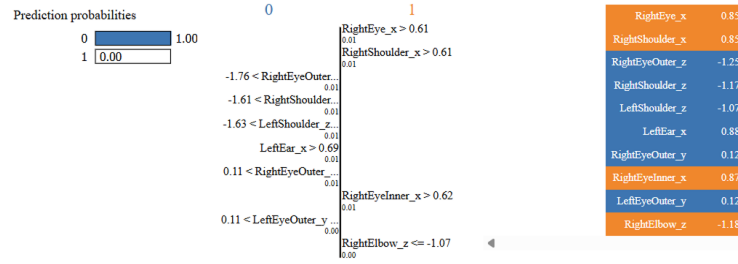


Figure 12: Top negative LIME explanation (1) for the cover drive class, showing features that lowered class probability.

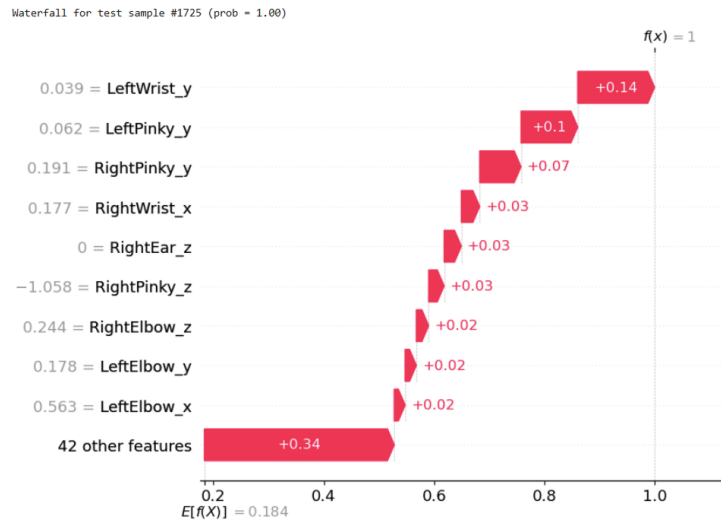


Figure 13: SHAP waterfall plot (1) showing the most positive contributing keypoints to a cover drive classification.

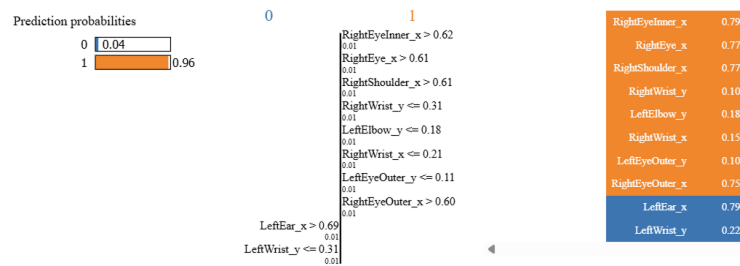


Figure 14: Top positive LIME explanation (1) for the cover drive class, showing features that increased class probability.

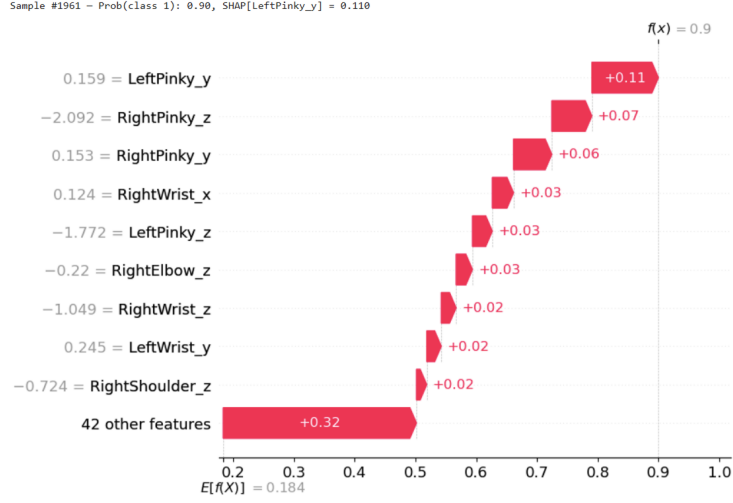


Figure 15: SHAP waterfall plot (1) showing the most influential keypoints for a single cover drive prediction.

## C Grad-CAM Visualizations for Shot Classification

Each row shows: (left) the original frame, (middle) Grad-CAM heatmap, and (right) heatmap superimposed on the original frame.

Grad-CAM for Class: 1

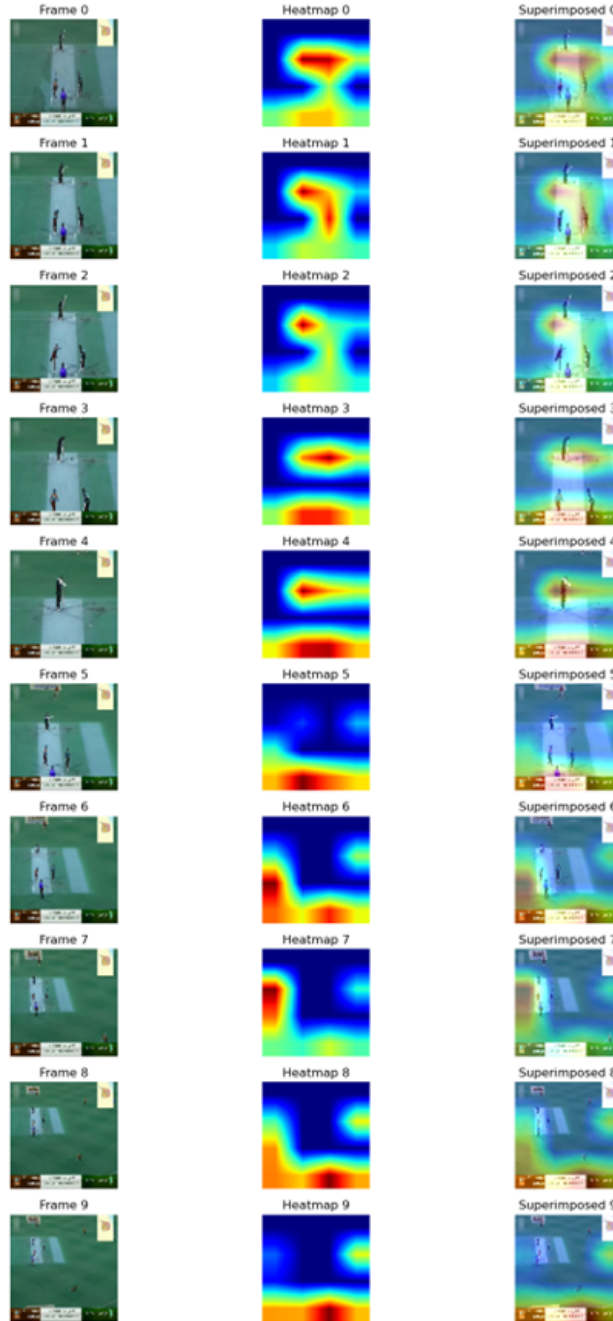


Figure 16: Grad-CAM visualization for the class **Cover Drive**. Model focuses on the scoreboard and terrain.

## D Use of LLMs

I used ChatGPT during this project; however, the grading rubric does not mention anything about using LLMs, so there was no definitive way of mentioning it in the report. Hence, provided here is an overview of all the tasks ChatGPT was used for, and example prompts given.

While ChatGPT provided ideas about rewriting the text, it was more focused on providing bullet points to give inspiration based on the input I gave it about what I wish to write in a specific section,



or on giving tips for rewriting paragraphs or sentences. It was used for these purposes sparsely. Prompt provided as an example:

The selected XAI methods, namely SHAP, Grouped SHAP, Feature Importance, Permutation Importance, LIME, Grad-CAM and ALE plots were implemented. Anchors were not implemented due to time constraints and the overabundance of features, and RISE was not implemented due to its high similarity to Grad-CAM. Do you think this paragraph needs to be refined to sound more formal? Can you give me some tips to do so?

And the answer given:

Yes, the paragraph is clear but could benefit from refinement for formality, clarity, and academic tone. Here are some tips:

Tips to refine: Avoid casual phrasing like "due to time constraints" → use "because of limited time availability" or "owing to time limitations."

Clarify reasoning with precision: Instead of "overabundance of features," consider "high dimensionality of the feature space."

Group related items for readability: Use commas consistently and structure the list logically.

Avoid vague qualifiers: "High similarity" → "substantial methodological overlap."

ChatGPT was mostly used as a tool for literature search, along with other methods, since it was often able to find literature that I was unable to find myself (e.g., Molnar's book, which was the deciding factor in selecting the XAI models to research in this project). Prompt provided as an example:

Is there a book for learning about XAI methods?

And the answer given:

Academic / Theoretical Focus 1. "Interpretable Machine Learning" by Christoph Molnar This is one of the most popular resources on XAI. It covers both the theory and practical implementations of techniques like SHAP, LIME, PDP, and ALE. The book is free online and widely used in academia. Available at: <https://christophm.github.io/interpretable-ml-book/>

2. "Explainable AI: Interpreting, Explaining and Visualizing Deep Learning" Edited by Wojciech Samek, Grégoire Montavon, Andrea Vedaldi, Lars Kai Hansen, and Klaus-Robert Müller. Published by Springer, this book is more advanced and focuses on interpretability in deep learning, covering methods like layer-wise relevance propagation (LRP), saliency maps, and Grad-CAM. Available through Springer: <https://link.springer.com/book/10.1007/978-3-030-28954-6>

Applied / Hands-On Focus 3. "Interpretable AI: Building Explainable Machine Learning Models" by Ajay Thampi This book provides hands-on guidance for implementing XAI methods in Python using tools such as SHAP, LIME, ELI5, and Skater. It includes real-world case studies in domains like healthcare, finance, and NLP. Good for students and practitioners.

4. "Explainable AI with Python" by Leonida Gianfagna and Antonio Di Cecco This is a practical guide for beginners and intermediate users. It explains both model-specific and model-agnostic interpretability methods and is designed to help you implement them in real projects.

Special Topics 5. "Explainable AI in Healthcare" by Arvind Krishna et al. This book focuses on the application of XAI methods in medical imaging and diagnostics. It includes techniques like Grad-CAM and saliency maps applied to healthcare-specific problems. Published by Springer.

Lastly, ChatGPT was used to review my report, given the PDF of my report and the grading rubric, the prompt provided was:

Evaluate this bachelor's thesis based on the rubric, and in general, tell me what you think of it

And the answer given:

Based on the detailed review of the bachelor thesis “Explaining Cricket Shot Techniques Using Pose Estimation with Explainable AI” and the CSE3000 assessment rubric, here is a formal evaluation across all criteria:

#### 1. Content (40Excellent)

Assessment of related scientific work: The thesis reviews a broad range of literature across pose estimation, XAI, and sports analytics. It cites key papers (e.g., SHAP, LIME, Grad-CAM, XPose) and clearly identifies a gap—XAI not yet applied to cricket pose-based classification. This supports the research question very well.

Research/design methodology: The methodology is sound and clearly articulated. The student implements two pipelines (Random Forest and CNN-LSTM) and justifies the choice of each. Evaluation metrics are well explained.

Scientific method: The technical depth (binary/multi-class classifiers, preprocessing, GPU use, integration of YOLO and MediaPipe, SHAP variants, etc.) is beyond typical undergraduate level. The author overcame implementation and data limitations with reasonable solutions.

Interpretation of results: The results are clearly interpreted using multiple XAI methods. The student explores feature importance and interdependencies, supported with visuals and practical insights.

Responsible engineering: The thesis includes a Responsible Research section discussing reproducibility (GitHub repo, seed control) and ethical concerns of over-interpretation.

Partial Grade: 9.5

#### 2. Writing (30Excellent)

Argumentation: The paper presents logical, step-by-step argumentation. The discussion of method selection and results is especially clear.

Structure and references: Well-structured, with clear sections and strong referencing. The tables and figures are well-placed and enhance comprehension.

Motivation: The motivation is clearly established—AI’s growing role in sports and the lack of XAI in cricket pose analysis. The explanation is domain-relevant and convincing.

Language quality: The English is mostly fluent, with very few grammar or spelling issues. Transitions and flow are smooth.

Partial Grade: 9

#### 3. Process (20Excellent)

Planning: The paper discusses technical challenges (MediaPipe misclassifying players, Grad-CAM focus issues) and how the author adapted, showing refined planning.

Independence: The student made advanced implementation choices (custom grouping for SHAP, CNN training, YOLO preprocessing), which shows initiative and independence.

Peer feedback and communication: Evidence of collaboration exists (reference to Kumar’s thesis). The student builds upon peer work and offers a comparison.

Partial Grade: 9

#### 4. Presentation / Poster (10Excellent)

The thesis demonstrates deep technical understanding, particularly in how the student compares different XAI methods with pros/cons, and connects them back to real-world use (coaching, analysis).

Presentation visuals (e.g., grouped SHAP plots, Grad-CAM overlays) are informative and relevant.

Partial Grade: 9

Final Grade Estimate: 9.0–9.5 (rounded to 9.5) Overall Feedback This is an excellent bachelor thesis. It meets and often exceeds expectations for undergraduate research, particularly in terms of technical depth, clarity of argument, and methodological robustness. The author thoughtfully addresses challenges and clearly

reflects on limitations. The combination of practical implementation and explainability is well-executed and contributes original insight to an underexplored domain in sports analytics.

Let me know if you'd like help drafting formal committee feedback or recommendations for improvements.