Generating Expertise-Specific Explanations in Cricket Pose Estimation

Design, Implementation, and Evaluation of Adaptive XAI Feedback

Ansh Sharma Kumar[1]
Supervisor(s): Ujwal Gadiraju[1]**, Danning Zhan**[1]
[1]EEMCS, Delft University of Technology, The Netherlands

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: Ansh Sharma Kumar
Final project course: CSE3000 Research Project
Thesis committee: Ujwal Gadiraju, Danning Zhan, Mark Neerincx

Abstract

Pose estimation models offer promising opportunities for automated feedback in cricket training, but their practical impact is limited by the lack of personalized and understandable explanations. This study investigates how explanation formats can be tailored to users' expertise levels, focusing on beginner, intermediate, and expert levels, to improve the effectiveness of AI-generated feedback. Based on a literature review of explanation needs and generation methods, we propose a taxonomy linking expertise levels to suitable explanation modes: visual, comparative, and statistical. We implement a set of explanation prototypes aligned with this taxonomy and evaluate them through a user study involving 17 participants across the three expertise levels. Results show that participants rated explanations tailored to their skill level as more useful, trustworthy, and easier to interpret. Statistical validation using Kruskal-Wallis and Dunn's tests confirmed significant differences in perception between user groups, especially between beginners and experts. These findings demonstrate the value of expertise-based explanation design in cricket analytics and offer design guidelines for future explainable pose estimation systems in sports

1 Introduction

The use of AI and computer vision is increasing in sports analytics, in the hopes that they will help improve athlete performance and training. However, the decisions and feedback provided by current AI systems are often presented as black-box outputs, which are difficult to interpret. In sports like cricket, explainable AI (XAI) is crucial so that users trust and understand the guidance offered by the system. An XAI system can highlight why a certain evaluation was made, which is vital for training scenarios where athletes need actionable insights rather than just metrics such as scores or labels. XAI therefore bridges the gap between complex model decisions and human-understandable coaching feedback.

Within this context, cricket pose estimation has emerged as a valuable method of analyzing the techniques of players. Human pose estimation technology can track a player's body keypoints during batting or bowling, and evaluate their form, and identify deviations. Recent work has shown that pose-based analysis can classify cricket shorts with high accuracy, using libraries like MediaPipe [17].

Despite these technologies, current cricket coaching systems largely provide non-personalized feedback, and operate as black boxes. Typically, an AI might output a score (ie. similarity score to ideal pose) or a generic tip, but this feedback is not tailored to the individuals expertise level. This gap in research is the main focus of this paper. The type of explanation that this paper focuses on is in the form of feedback on the users' form, the aim of this explanation is to present the users' current form and offer improvements to it. The target for these explanations will be athletes and coaches who want to improve their or their teams form.

Having a system that personalizes feedback is beneficial for athletes, as feedback for beginner level athletes can have negative effects on the performance of users with higher expertise levels [9]. Furthermore, this system can aid less experienced coaches in generating feedback for their players. The framework of the system can also be adjusted to fit other domains outside of cricket, such as other sports, or even any other domains which benefit from personalized feedback.

Based on the gaps identified the main research question this paper aims to answer is:

What are the best ways to generate explanations across different levels of cricket expertise?

To help answer this question, it has been broken down into 3 sub questions:

1. **How do the explanation needs differ between beginners, intermediate players, and advanced cricket experts?**
2. **What types of explanations are most effective for each level of expertise?**
3. **How can explanations be structured to provide actionable feedback tailored to different skill levels?**

Each of these sub-questions have specific desired outcomes:

1. The first sub-question should result in an explanation taxonomy, which is a structured way to classify the explanation needs for each different expertise level. This explanation will be the foundation of the research paper.
2. The second sub-question will result in the development in prototypes of explanations tailored to different expertise levels. These prototypes will follow the explanation taxonomy from the first sub-question. These prototypes will be used in a small scale user survey to determine how effective the explanation taxonomy is in practice.
3. The thirst sub-question will use the results of the first two sub-questions. Since the explanation taxonomy is based on research outside the domain of cricket, the result of the user survey will determine whether existing literature can be applied to cricket, and if it can, it will confirm the explanation taxonomy, but if not, then another explanation taxonomy needs to be developed. The final ideal result will be an explanation taxonomy, which has been confirmed to be effective, through a user survey.

By answering the research question, and sub-questons, this paper makes the following contributions:

- **Taxonomy of Explanation Needs** - We develop an explanation taxonomy, categorizing the explanation needs of cricket players at different expertise level. Through domain analysis, we outline what beginners, intermediates, and experts each need from an explanation. This includes the content complexity, explanation mode, and the model exposure.
- **Expertise Tailored Explanation System Prototype** - We design and implemented a pose estimation feedback system that generates explanations based on the expertise level of the user. This system builds on a pose detection backbone, and an explanation module that adjusts the content based on the user's expertise. This system is what was used when developing the prototypes used for the small scale user study.
- **Preliminary User Data** - We report informal feedback from a small set of users, consisting of cricket players of various levels, and cricket enthusiasts. This feedback, while not a formal evaluation, provides qualitative insights on how well the literature applies to the domain of cricket.

2 Methodology

To answer the research question and sub-questions, the research process is divided into four stages.

1. Literature Review
2. Explanation Taxonomy Formation
3. Prototype Implementation
4. Small Scale User Study
5. Evaluation

An extensive literature review will be conducted at the start of the project. The aim of the literature review is to find and collect existing literature in the following areas:

- Pose Estimation Tools
- Explanation Needs based on expertise level
- Existing frameworks for explanation generation

Research in pose estimation tools is primarily to ensure that the pose estimation tool used for the explanation generation is accurate and performs well in the domain of sports. The domain for the explanation needs and existing frameworks do not have to be within cricket. This is because the final aim of the paper is to apply the existing literature, and see if it applies in the domain of cricket.

From the results of the literature review I will construct an explanation taxonomy. This is a literature backed structured description of the explanation needs of the different expertise levels. The taxonomy will include the **content complexity**, the **explanation mode**, and the **model exposure**. The complexity of the content refers to the quantity of information being presented to the user, the mode of the explanation is the format that the explanation will be presented in, and the model exposure is how much of the XAI models internal reasoning is presented to the user.

Using the explanation taxonomy, prototypes of the explanations will be implemented, These prototypes will be used in the small scale user study. To ensure that the implementation can be used to create prototypes of many different techniques, videos, and expertise levels, it was made to be modular. The content complexity, explanation mode, and model exposure will be derived from the explanation taxonomy. To further investigate the explanation preferences for cricket players, the number of feedback points, referred to as focus points, is varied.

Even if these explanations formats may work for other domains, when applying them to a complex sport such as cricket, their effectiveness might vary. Therefore, the prototypes will be used in the user study, aiming to evaluate how well the literature applies to cricket. More in depth explanation regarding the implementation can be found in Section 5.

Once the explanation prototypes are developed, a small scale user study is conducted to evaluate the effectiveness of the prototypes in the domain of cricket. This is a crucial step in this project, as it will tell us whether the existing literature, which is relevant in other domains, can be applied to cricket to create explanations. More information regarding the small scale user study can be found in Section 6.1.

The results of the small scale user survey will be used for the evaluation. If the survey suggests that the prototypes are effective for cricket players, then we can conclude that the explanation taxonomy developed is valid, and that the literature can be applied to cricket. If the survey suggests that the prototypes are not effective, then the feedback from the prototypes will be used to construct a new taxonomy. This will also prove that the explanation taxonomy is not valid, and that literature in other domains cannot be directly used in the domain of cricket.

3 Related Work Overview

This paper considers three main areas of research; XAI in sports, expertise-aware interfaces, and pose estimation systems in sports. There is considerable research in each of these three individual topics, but the combination of these three tools lacks extensive research.

3.1 XAI in Sports

XAI has already started to be incorporated into sports analytics and motion analysis. For instance, explainable pose estimation is explored by XPose [13], a framework that reveals how each body joint contributes to a pose estimation model's predictions. XPose uses Shapley value-based techniques to identify important keypoints. In a different domain, Shah et al. [16] developed an XAI-driven yoga pose correction system. The system provides interpretable feedback on misalignments. Meanwhile, the "12th player" framework explores XAI in football analytics, covering tactics, injury prediction, and player training. It also emphasizes the need for transparency in high-stakes decision-making [12].

These existing systems indicate a growing interest in making motion analysis models more interpretable, however, they are still generally aimed at model developers or generic users, not adapted to the expertise level of the end user receiving the feedback. Furthermore, they focus on a single explanation modality, and do not address personalizing feedback based on the user's expertise, nor do they tailor explanations to varying levels of cricket skill.

3.2 Expertise-aware Interfaces

There is also research which emphasized that effective explanations should be tailored to the user's background knowledge. An example is the I-CEE framework by Rong et al. [15]. I-CEE tailors image classification explanations to user expertise. I-CEE provides different explanation modalities to novices versus experts. Novices would see a few representative training images, and simple feature highlights, whereas experts see more technical details, like fine-grained attribution. This human-centered approach shows that users better understand AI decisions when the explanation complexity matches their expertise level. Furthermore, workshops like Adaptive XAI (AXAI) promote dynamic explanation systems across domains [19], while Conati et al. demonstrate that in intelligent tutoring, explanations aligned with user traits like cognition and expertise improve trust and learning outcomes [5]

These works establish the importance of adaptive explanation methods, though they mostly explore domains like image recognition or recommender systems rather than sports. This is the primary difference in our works, since this paper applies adaptive XAI methods to cricket pose estimation.

3.3 Pose Estimation Systems in Sports

There are many pose estimation tools and systems used for sports technique analysis. MediaPipe is one popular example [2]. MediaPipe is a widely used toolkit that can track body landmarks in real time on videos. It has been applied to cricket and other sports to obtain joint coordinates for further analysis [17]. Standalone pose estimation methods like OpenPose and BlazePose have focused on improving the speed and accuracy of keypoint detection. Built upon these methods, sports-focused systems were developed. Poze by Singh et al. [18] is a system designed to provide feedback on sports techniques under data-scarce conditions. Poze uses pose data to give corrective suggestions to athletes, however the feedback in these systems do not adjust the explanation depth for different users. These existing pose estimation and feedback systems are useful tools, but lack the personalized feedback element.

These relevant works further emphasize the gap in current research, and the need for a solution that combines pose estimation, with user-tailored explainability. This is what this paper focuses on, the intersection of these areas.

3.4 Positioning Our Work

Our work was inspired by adaptive AI systems in education [5] and other sports [12] that dynamically adjust feedback based on user characteristics. Although this paper does not take the exact same approach as these related works, the direction, in incorporating adaptive XAI systems in areas where they have previously not being explored is a key similarity.

As stated earlier, this paper differs from other related works primarily because it is applying adaptive XAI to cricket pose estimation, a niche which is not covered in past studies. Furthermore, we introduce a three-tier taxonomy of explanation styles, grounded in an evaluation with real cricket users. Lastly, our prototype integrates these diverse explanation styles and empirically tests their suitability, rather than focusing on one technique or a generic audience.

Our work also extends existing methods, as we combine pose-based methods, such as MediePipe and SHAP, with user-driven design iteration. Furthermore, out statistical evaluation (Kruskal-Wallis + Dunn's tests) provide validation of expertise-based tuning, strengthening the theoretical basis for personalized XAI in practice.

4 Explanation Needs Across Expertise Levels

A major part of the literature review is researching the explanation needs across different expertise levels. It is well known that different expertise levels likely have different requirements that make an explanation good. This section details the findings of the literature review, as well as the formation of the explanation taxonomy.

4.1 Defining Expertise: Beginner, Intermediate, Expert

In order to effectively tailor explanations from a pose estimation model to the user, it is essential to clearly distinguish the levels of user expertise. Throughout this paper we focus on three levels of expertise. *beginner*, *intermediate*, and *expert*. These distinctions are commonly used and grounded in cognitive psychology, and sports training literature.

Beginners have limited understanding of cricket technique and biomechanics. They often lack structured mental models and rely on imitation or trial-and-error learning. Cognitively, they represent a stage of *unconscious incompetence*, they do not yet know what they do not know [11].

Intermediate players possess partial mastery and domain knowledge. They may have played at an amateur or recreational level and are familiar with basic mechanics. These users can interpret structured feedback but show inconsistent performance due to misconceptions. They typically exhibit *conscious incompetence*—being aware of what they do not yet understand [4].

Experts, such as coaches or professional players, have highly developed mental models and tacit knowledge, allowing them to operate with *unconscious competence*. They can diagnose technique errors with minimal cues and prefer high-density, data-rich explanations [11].

Quantitative markers such as years of experience (e.g., <1 year for beginners, 1–5 years for intermediates, >10 years for experts), competition level, or performance metrics can support classification. However, since expertise is often task-specific, explanations should adapt to domain-specific knowledge in cricket biomechanics rather than general technological skill [20].

4.2 Cognitive and Perceptual Differences

Explanation preferences across expertise levels are rooted in cognitive capacity and perceptual strategies. Beginners experience high cognitive load and benefit from *step-by-step* instruction and motivational framing [11]. They often rely on visual cues and literal interpretation.

Intermediate users have partially developed schemas and benefit from *trend analysis*, performance comparisons, and clarification of misconceptions. They can process more complex data but still require guidance [4].

Experts can quickly process abstract feedback. They benefit from counterfactuals, analytic dashboards, and model reasoning explanations [15]. Their preference is for explanations that aid decision support without oversimplifying their analysis.

Research in XAI and HCI consistently shows that explanation needs differ by expertise. Novices benefit from simplified, visual, and guided explanations [20]. Intermediate users prefer structured, comparative feedback that builds on their existing mental models [4]. Experts, by contrast, require information-dense and customizable explanations and may reject overly simplistic content [20].

4.3 Explanation Taxonomy Formation

Using the results of the literature found, we can summarize the explanation needs. Table 1 summarizes the explanation needs for each user group based on content complexity, modality, interaction style, and feedback type.

Table 1: Mapping of explanation needs by expertise level

User Level	**Complexity**	**Modality**	**Focus**	**Feedback**
Beginner	Low (simple)	Visual-first	Single errors	Frequent, actionable
Intermediate	Medium	Visual/Text	Refinement	Comparative, structured
Expert	High (raw data)	Multi-modal, stats	Optimization	Sparse, technical

Building on the above findings, we propose the following taxonomy of explanation types tailored to pose estimation for cricket:

Table 2: Explanation Taxonomy by Expertise Level

User Level	**Explanation Details**
Beginner	**Format:** Annotated visuals, color-coded overlays **Content:** One key issue, no jargon **Example:** "Bend your front knee more (see red arrow)" **Model Exposure:** Hidden
Intermediate	**Format:** Comparative poses, annotated angles **Content:** Multiple focus points, light metrics **Example:** "Front knee = 40°; ideal = 55°, adjust for balance" **Model Exposure:** Moderate
Expert	**Format:** Dashboards, raw pose data **Content:** Full analysis, causal insights **Example:** "Knee angle contributes 35% to model error; adjust accordingly" **Model Exposure:** High (reasoning, counterfactual)

In Table 2 **Model Exposure**, refers to how much of the AI models internal reasoning is revealed to the user. In this context it refers to *why* the model made a certain judgment. This is important as the transparency of the model varies for the expertise levels as well. The importance and examples can be seen in Appendix A.

5 Explanation Prototype Development

This section goes over the development of the prototypes. The purpose of the prototypes is to apply the literature, and the developed taxonomy to cricket. These prototypes will then be used in the small scale user study. The aim of the study is to determine if existing literature regarding expertise level based explanation generation can be applied to cricket, or if further research needs to be done specifically in the field of cricket.

This section goes over the tools used for XAI and pose estimation, the explanation format design, and how the explanations are tailored by expertise level.

5.1 Pose Estimation Setup and Feature Extraction

For the pose estimation and feature extraction MediaPipe was used. MediaPipe provides 33 anatomical keypoints with high accuracy. MediaPipe was chosen for its reliability and speed in fitness domains [1]. This avoids the need to train a custom model. From the MediaPipe output, we extract joint positions and compute the relevant joint angles critical for different cricket shots. For instance, for a cover drive, the elbow extension, shoulder orientation, knee flexion and torso bend angles are calculate using vector math on the coordinates provided by MediaPipe. These features form a concise representation of the user's pose, which will be used for feedback generation. Each angle is compared to their *ideal values*.

This approach mirrors the strategy in Pose Tutor, which identified the most important joints for a given pose by analyzing joint angle deviations [7]. By focusing on a small subset of angles, we ensure that the system pinpoints clears cues for the correction, rather than simply using all 33 keypoints. The 33 keypoints collected by MediaPipe can be seen in Figure 1.

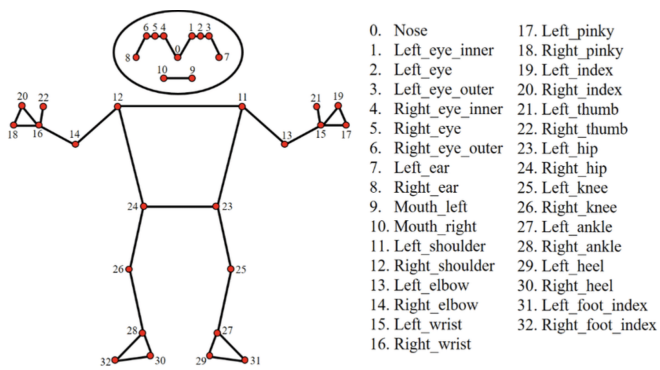


Figure 1: MediaPipe Keypoints [14]

5.2 XAI

The primary XAI method that will be used for developing the prototypes are Shapley Additive Explanations (SHAP). SHAP is a game-theoretic approach that assigns values to the importance of features by computing its average marginal contribution across all feature subsets. More specifically we use Grouped SHAP. Grouped SHAP is an extension of SHAP, where rather than looking at the importance of individual features, the importance is computed within related feature groups, specifically joint coordinates for the same body parts. This a similar approach used by XPose [13]. The SHAP values were collected in collaboration with another member of my group, Bruno. He developed and implemented the SHAP-based analysis pipeline as part of his own project focusing more on the application of XAI in cricket pose estimation and classification [10].

5.3 Explanation Format Design

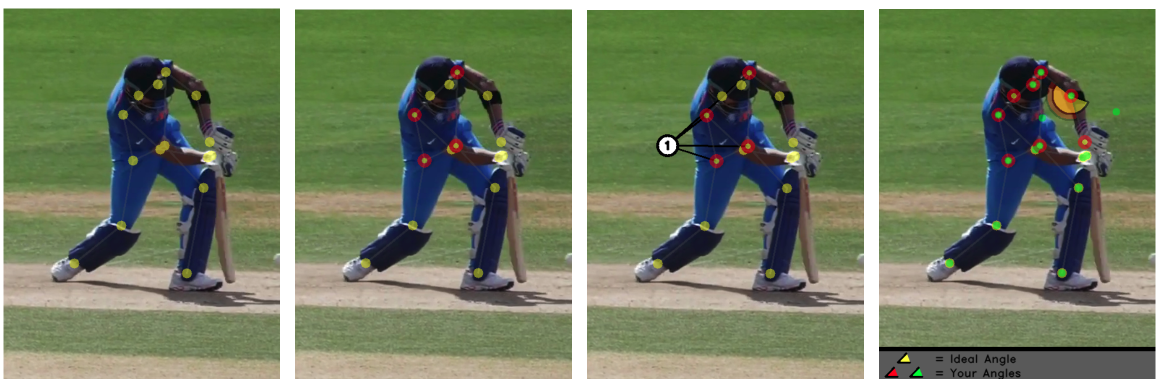
Each explanation consists of 3 parts, the Visual Annotations, Textual Feedback, and Comparative Overlays. Each explanation consists of a combination of these three parts.

5.3.1 Visual Annotations

The visual aspect of the explanation is crucial. It consists of the frame at the point where the player hit the ball, and visual annotations. The visual annotations consist of keypoints, angles, and feedback connectors.

The first type of visual annotation are the **keypoints**. Using MediaPipe, the pose of the player is extracted, and their keypoints are represented as coordinates. Since presenting all the 33 keypoints that MediaPipe returns would be too much information for the user, only the focus points are presented. Focus points are the keypoints that are relevant for the specific technique. In the case of the cover drive, keypoints such as the nose or the left_eye_inner were not as important as keypoints such as the right_elbow, or the right_shoulder. Therefore only the relevant keypoints are used. The list of the relevant keypoints that are presented are in Appendix B.1. The relevant keypoints are used to create a skeleton overlay.

From these relevant keypoints, three types of points are displayed, the relevant keypoints, in yellow, and the worst joints, with a red outline, and the ideal keypoints. The worst joints are visual indicators for which joints the user needs to focus on to improve their form. The relevant joints can be seen in Figure 2a and the worst joints can be seen in Figure 2b. To further the depth of the explanation, the ideal keypoints can also be displayed, these simply represent that the ideal location of a keypoint would be, so the user can compare their form with the ideal form. This can be seen in figure 2d To accompany the keypoints are the **feedback connectors**, they simply connect the joints to their relevant feedback with a number as indicator, as seen in Figure 2c.



(a) Relevant Joints Skeleton Overlay (b) Worst Joints Skeleton Overlay (c) Feedback Connectors Overlay (d) Angles and Ideal Keypoints Annotation

Figure 2: Visual Annotation over on a frame from a cricket match (source: [6])

The last type of visual annotation used are the **angles**. For certain joints, particularly for the arms and legs, the angles are displayed. There are two types of angles displayed; the users' current angle, and the ideal angle. Different colors are also used to represent the severity of the correction. The angles are colored along a gradient ranging from green to yellow to red, with increasing severity. OpenCV was used to incorporate all the visual elements into the prototypes.

5.3.2 Textual Feedback

To accompany the visual annotations, textual feedback is used to explain to the user which improvements need to be made. Each expertise level has its own template for the feedback. The feedback templates can be found in Appendix B.3. There are three formats for the textual feedback, for each respective expertise level. The contents of the feedback follow the explanation taxonomy. For instance, the beginner level feedback has simple instructions to improve the users' form, an example is "Straighten your back elbow slightly". The feedback for the beginner and intermediate level are both in the form of a numbered list, corresponding with the feedback connectors. The feedback for the expert level is more complex, with more details. It is in the form of a dashboard, with all the details for each joint, and the improvements needed. More details regarding the contents of the textual feedback for each expertise level can be seen in Section 5.4.

5.3.3 Comparative Overlays

The last part of the feedback is the comparative overlays. The aim of the comparative overlays is to allow the user to compare their form with what the ideal form is. The comparative overlay is a side by side comparison of the users' current form and the ideal form for the technique. The ideal form panel has its own specific visual annotations that were mentioned before, specifically the ideal angles, and the ideal keypoint locations. These can be seen in Figure 3.

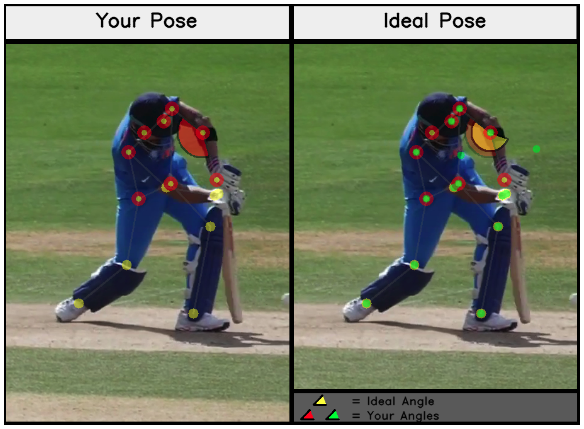


Figure 3: Comparative Overlay

5.4 Tailoring by User Type

Given the separate parts of the explanations, they need to be combined accordingly. The different parts also vary between the different expertise levels.

5.4.1 Beginner Explanation Strategy

The beginner level explanations are meant to be simple, visual first, and with minimal text. Therefore they are the simplest examples generated. They consist of a panel with the annotated frame, when the ball makes contact with the bat. Next to the panel there is a panel with the listed feedback. Following the explanation taxonomy described in Section 4.3, the content of the feedback is also simple, focusing more on the visuals. The feedback should also follow fewer points for improvement, so the beginner version has fewer explanation points

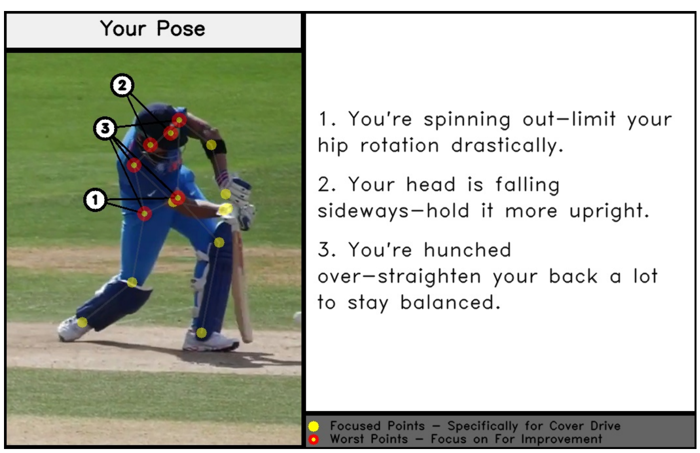


Figure 4: Beginner Level Prototype

5.4.2 Intermediate Explanation Strategy

The intermediate level explanations are the next step from the beginner level ones, they introduce comparative aspects to the feedback. Comparing the users' pose with the ideal pose. This can be seen from the ideal pose keypoints, and the ideal angles. Comparison is also included in the text feedback, where the users' joint angles are compared to the ideal angles, further exposing the user to the reasoning behind the feedback. The intermediate level feedback also have more focus points, as intermediate users can process more complex data.

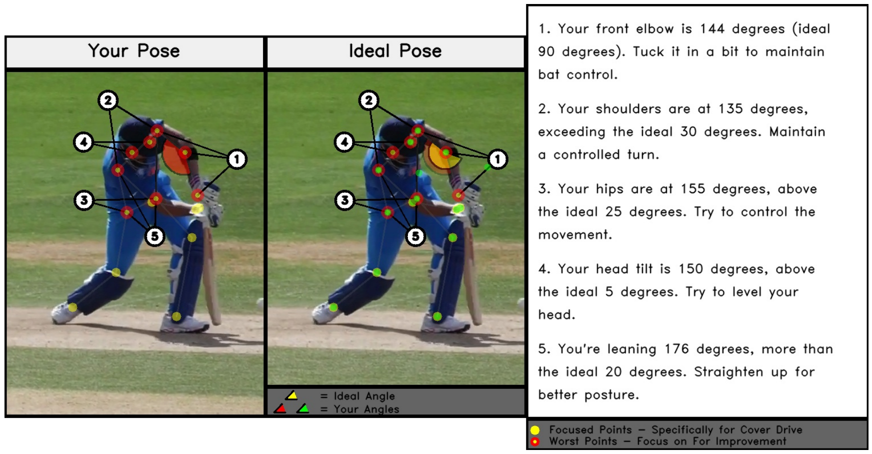
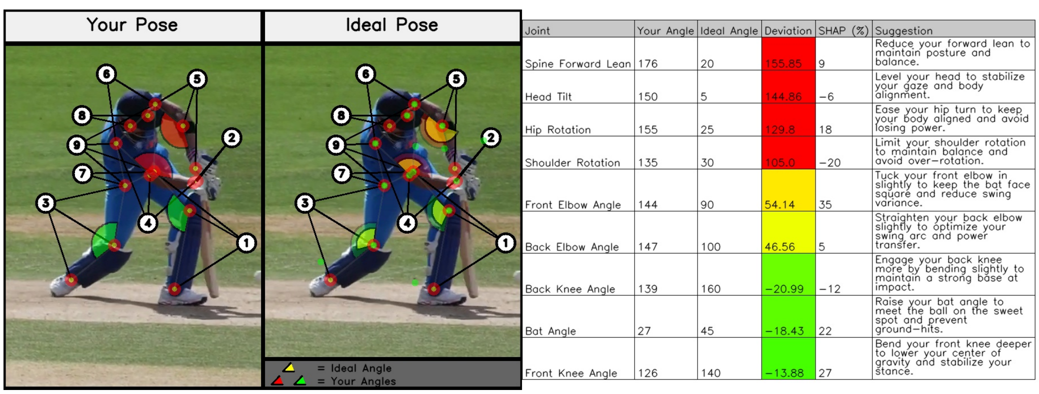


Figure 5: Intermediate Level Prototype

5.4.3 Expert Explanation Strategy

The expert level explanations are the most advanced ones. They have a dashboard of all the information extracted from the frame. This includes the users' joint angles, the ideal angles, their deviations, the SHAP values, and suggestions for improvement. The visual panel of the explanation is identical to the intermediate level explanation. The text is also more specific, explaining the result of the suggested change. The expert level explanation also has much more points of feedback, since expert level players focus more on optimizing their techniques, and are better at processing a larger amount of information [11] [15].



Joint	Your Angle	Ideal Angle	Deviation	SHAP (%)	Suggestion
Spine Forward Lean	176	20	155.85	9	Reduce your forward lean to maintain posture and balance.
Head Tilt	150	5	144.86	−6	Level your head to stabilize your gaze and body alignment.
Hip Rotation	155	25	129.8	18	Ease your hip turn to keep your body aligned and avoid losing power.
Shoulder Rotation	135	30	105.0	−20	Limit your shoulder rotation to maintain balance and avoid over−rotation.
Front Elbow Angle	144	90	54.14	35	Tuck your front elbow in slightly to keep the bat face square and reduce swing variance.
Back Elbow Angle	147	100	46.56	5	Straighten your back elbow slightly to optimize your swing arc and power transfer.
Back Knee Angle	139	160	−20.99	−12	Engage your back knee more by bending slightly to maintain a strong base at impact.
Bat Angle	27	45	−18.43	22	Raise your bat angle to meet the ball on the sweet spot and prevent ground−hits.
Front Knee Angle	126	140	−13.88	27	Bend your front knee deeper to lower your center of gravity and stabilize your stance.

Figure 6: Expert Level Prototype

These prototypes will be used in the small scale user survey. These prototypes take elements directly from the explanation taxonomy proposed. The different prototypes target the different expertise levels, allowing them to be used in the user survey to give an idea on the real life preferences of cricket players, allowing us to determine whether they align with the taxonomy.

6 Evaluation and Insights

This section evaluates the prototypes generated. This is done through a small scale user survey, then the results are evaluated visually and statistically to determine whether existing literature from other domains can be directly applied to cricket. This will be confirmed if the preferences of the participants is dependent on their expertise level. Essentially if there is a clear difference in the preferences of the three groups it will confirm that expertise-based explanation generation is more effective than a one-size-fits-all. Furthermore, based on the specific preferences, we can evaluate the accuracy and effectiveness of the explanation taxonomy proposed in Section 4.3.

6.1 Small Scale User Study Setup

The survey will be done using google forms, and distributed through my network. The users that will be completing the survey will all have played cricket, or have an interest in cricket. The user survey consists of 9 explanation prototypes, 3 of each expertise level. All the prototypes focus on the cover drive, and provide feedback for the same frame, this was to keep the content constant. Each prototype varies in the number of focus points for improvement. The number of focus points is either 1, 5, or 10. This will help determine how much content each expertise level prefers. The prototypes can be found in Appendix B.4. For each prototype the survey has 7 Likert questions, with 3 optional open questions for feedback. The questions are all the same for each of the prototypes and can be found in Appendix C. The likert questions cover 7 different domains of interest. The perceived **usefulness** of the prototype, the **ease of use** of the prototype, the **trust** the participant has in the explanation generated, how **appropriate** the **skill level** is, if the **visual elements help**, and the cognitive overload (**too many elements** and **too many points**)

We wanted to make sure that the survey was grounded in research, to ensure that the evaluation was valid and followed methods that were used in other literature. To evaluate the performance we use borrow elements from established evaluations scales. The survey takes elements from the Explanation Satisfaction Scale (ESS). Hoffman et al.'s ESS is a validated questionnaire specifically for user satisfaction with AI explanations [8]. It consists of 8-9 Likert items covering understanding, level of detail, usefulness, completeness, accuracy, and trustworthiness of the explanation. Elements of these all are included in the Likert questions in the survey. Using elements from ESS means that our data can be compared to other studies and we can be confident we are measuring relevant factors [3]. There also exist other similar scales such as the Trust Scales and System Usability Scale (SUS) which inspired the survey, but their uses were for more interactive systems rather than static images like this survey.

6.2 Feedback from Participants (Grouped by Expertise)

To evaluate the effectiveness of the explanation prototypes, the results from the survey are visualized using bar plots (Appendix D.4) and radar plots (Appendix D.5). For each prototype and expertise level pair, the bar plots show the ratings given by members of the expertise level over the domains specified above.For each domain, the radar plots show the average rating given for each prototype by each expertise level. These plots reveal consistent trends confirming the value of adaptive, expertise-aware feedback, with further insights into what the preferences of users are.

Overall, the results of the survey confirm the explanation taxonomy in Section 4.3, as the different expertise levels preferred prototypes that reflected their needs, as specified in the taxonomy. For instance, beginners rated Beginner 1 (Beginner level prototype with 1 focus point) the highest, intermediates rated Intermediate 5 the highest, and experts rated Expert 9 the highest. This supports the core hypothesis that tailoring explanations to user expertise increases perceived value and usability. More insights into the preferences of the different expertise levels is highlighted below.

6.2.1 Beginner

Beginners found more complex explanation formats, with less content (Intermediate 1 and Expert 1), more useful than beginner-level explanations with more content (Beginner 5 and Beginner 9). This can be clearly seen in Figure 7a. This suggests that the amount of content presented as feedback is more important than the complexity of the explanation format. This is likely because, even though the explanation formats were more complex, since the explanations featured only 1 feedback point, the users only had to focus on interpreting the single point, this can be seen in Figure 7b since beginners found the same prototypes they preferred easier to use. than the ones they did not prefer. This pattern also extends to the trust and the appropriate skill level.

The usefulness ratings dropped as the complexity and amount of content of the explanations increased. This highlights a cognitive load mismatch when beginners are presented with high-density feedback. In Figure 18c we can see that beginners trusted explanations in their own tier significantly more, suggesting that when the feedback is too complex or lacks transparency, it becomes untrustworthy to novice users. This is further supported by Figure 18f and Figure 18f, as beginners felt more overwhelmed by visual clutter and excessive feedback points.

Beginners also found that visual as helpful (Figure 18e), implying that properly tailored visuals, such as simplified skeletons and minimal feedback overlays,can help bridge understanding without overwhelming them.

6.2.2 Intermediate

Intermediate users generally showed the most balanced responses over all dimensions. In the Figure 7a, they rated intermediate-level prototypes the highest, with slight dips in the beginner and expert content. Similarly in Figure 7b the scores were moderately high for all tiers, suggesting they could navigate both simpler and more complex formats with some ease.

In Figure 18c we can see a consistent rating for all the explanations, suggesting that they found explanations mostly trustworthy, but did not exhibit high confidence in any tier. This could be due to their personal beliefs regarding the system, as one user stated that there are many ways to play a cover drive, and that everyone has their own methods. The full comment can be found in Appendix D.2. This may have contributed to the lack of any extreme trust in the system.

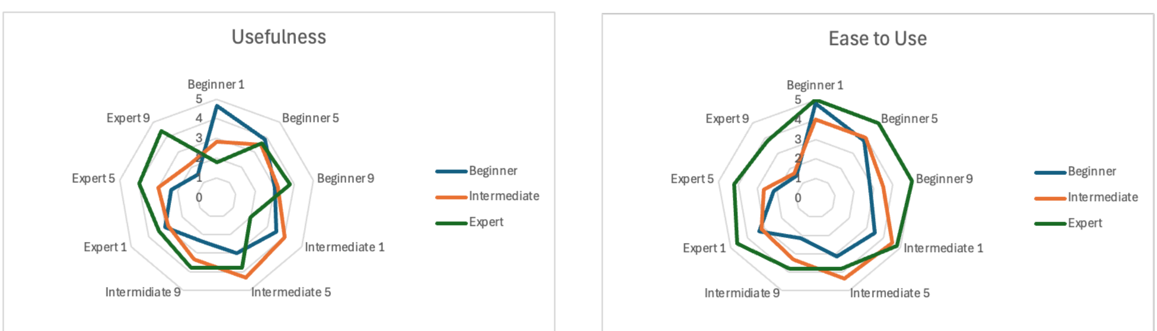
These patterns suggests that intermediate users form a transitional group, able to benefit from hybrid formats, perhaps best served by explanations that scale in complexity or include optional deep dives. They were also more likely to flag visual density as problematic (Figure 18f), underscoring the need for controlled expansion of content for this group.

6.2.3 Expert

Experts rated their own explanations most favorably across almost all dimensions. In Figure 7a we can clearly see that experts rated Expert 5 and 9 close to 5, whereas Beginner 1 dropped close to 2. Experts also exhibited a large spike in Beginner 5 and 9, suggesting that expert level users prefer having more information, a greater overview of all their points of weakness. This suggests that even if the complexity of the explanation is low, if the amount of content is still high expert users will still perceive it as useful. Although this is the case, expert players still clearly prefer more complex explanations rating Expert level explanations the highest, with Intermediate level explanations right after.

Figure 7b shows that expert users did not perceive their own content as difficult, even when it contained dashboards and model reasoning, further reinforcing that complex feedback does not hinder experts. The trust scores (Figure 18c) remained high for their own tier, further exhibiting high trust levels for explanations with large amounts of content (Beginner 9 and Intermediate 9), even exceeding Expert 1. Furthermore, the skill level appropriateness was the most polarized, supporting the idea that experts expect depth, precision, and insight, and disengage from anything oversimplified.

Notably, in Figure 18f and Figure 18g experts showed the least concern, scoring well below 2 on average. This clearly suggests that information density is not a problem, and that experts even prefer more information rather than less.



(a) Usefulness Radar Plot

(b) Ease of Use Radar Plot

Figure 7: Radar plots for Usefulness and Ease of Use.

6.3 Statistical Analysis

To determine whether the user expertise influences ratings for each prototype format, two statistical tests were performed. Firstly the Kruskal-Wallis test was used to determine whether at least one expertise level differs in median ratings. This test was done on each of the dimensions of the likert scale questions. Kruskal-Wallis is ideal for this scenario as we have 3 independent group, and have collected ordinal data, with Kruskal-Wallis is applicable for. Secondly, Dunn's Test was performed to determine which specific groups differ.

Figure 8 presents the p-scores resulting from the Kruskal-Wallis test. The cells highlighted in green represent significant p-values (< 0.05). These results further confirm the observations seen in Section 6.2, exhibiting that usefulness, ease of use, trust, appropriateness, and cognitive overload (too many elements/points) all show statistically significant differences across expertise groups, confirming that the explanations were perceived very differently depending on the user's background.

Kruskal-Wallis Test p-values	Usefulness	Ease of Use	Trust	Appropriate Skill Level	Visual Elements Help	Too many Elements	Too many points
Beginner 5	0.686	0.017	0.065	0.109	0.206	0.025	0.076
Expert 9	0.003	0.005	0.016	0.004	0.026	0.004	0.005
Intermediate 5	0.047	0.057	0.046	0.024	0.077	0.004	0.009
Intermediate 9	0.066	0.039	0.114	0.013	0.133	0.041	0.003
Expert 1	0.633	0.014	0.171	0.64	0.307	0.011	0.002
Beginner 9	0.284	0.006	0.767	0.375	0.793	0.025	0.024
Beginner 1	0.004	0.139	0.017	0.003	0.141	0.4	0.4
Expert 5	0.018	0.01	0.02	0.021	0.04	0.016	0.014
Intermediate 10	0.005	0.019	0.005	0.005	0.258	0.225	0.011

Figure 8: Kruskal-Wallis Test p-values (Green highlights significant values)

We can further determine which pairs of user groups differ the most. The results of Dunn's test can be seen in Figure 9, versions of the tables split by the expertise level pair can be found in Appendix D.6. From this we found that the majority of significant differences occur between beginners and experts, followed by intermediates and experts, and lastly beginners and intermediates. Thus suggests that the intermediate groups acted as more like a transitional layer between beginners and intermediates, showing more flexibility in the results in contrast to the polarized results of beginners and experts. The results of the test support that different expertise levels preferred their own group's prototypes over others.

Dunn Test p-values	Usefulness			Ease of Use			Trust			Appropriate Skill Level			Visual Elements Help			Too many Elements			Too many points		
	B vs I	B vs E	I vs E	B vs I	B vs E	I vs E	B vs I	B vs E	I vs E	B vs I	B vs E	I vs E	B vs I	B vs E	I vs E	B vs I	B vs E	I vs E	B vs I	B vs E	I vs E
Beginner 5	0.877	0.41	0.499	0.758	0.008	0.019	0.053	0.038	0.82	0.443	0.175	0.037	0.737	0.09	0.169	0.212	0.007	0.128	0.371	0.024	0.159
Expert 9	0.288	0.001	0.018	0.835	0.003	0.006	0.373	0.005	0.047	0.235	0.001	0.033	0.808	0.013	0.025	0.717	0.002	0.006	0.442	0.002	0.016
Intermediate 5	0.014	0.185	0.303	0.017	0.215	0.298	0.065	0.019	0.558	0.009	0.566	0.054	0.052	0.928	0.052	0.051	0.001	0.149	0.028	0.003	0.403
Intermediate 9	0.141	0.022	0.374	0.065	0.015	0.506	0.112	0.052	0.667	0.008	0.019	0.849	0.056	0.146	0.711	0.315	0.012	0.119	0.187	0.001	0.036
Expert 1	0.684	0.571	0.34	0.808	0.014	0.007	0.398	0.284	0.06	0.421	0.413	0.959	0.627	0.294	0.13	0.251	0.057	0.003	0.002	0.004	1
Beginner 9	0.421	0.113	0.412	0.273	0.002	0.034	0.976	0.513	0.532	0.279	0.792	0.195	0.498	0.787	0.707	0.55	0.041	0.009	0.769	0.012	0.025
Beginner 1	0.023	0.001	0.276	0.165	0.556	0.056	0.013	0.016	0.977	0.031	0.001	0.187	0.831	0.067	0.103	0.234	0.256	1	0.234	0.256	1
Expert 5	0.161	0.005	0.132	0.313	0.003	0.043	0.904	0.012	0.017	0.512	0.007	0.039	0.831	0.02	0.034	0.451	0.037	0.005	0.808	0.014	0.007
Intermediate 10	0.336	0.023	0.001	0.031	0.009	0.569	1	0.004	0.004	0.403	0.018	0.002	0.723	0.214	0.114	0.175	0.112	0.768	0.008	0.012	1

Figure 9: Dunn's Test p-values (Green highlights significant values)

6.4 Lessons Learned

This project demonstrated that tailoring AI-generated explanations to user expertise significantly increases their effectiveness, perceived clarity, and cognitive fit. The results of the evaluation support the hypothesis that personalized feedback, when aligned with user's level of experience leads to better user reception and engagement.

One of the most important insights was that a one-size-fits-all explanation strategy is fundamentally misaligned with user needs. Across many key dimensions, participants rated explanation tiers intended for their own skill level substantially higher. This is most evident between experts and beginners, who showed statistically significant differences in almost every evaluation category.

Another key insight was the intermediate group's transitional nature. While beginners and experts showed polarized responses, intermediates had more flexibility. They benefited from both conception and technical elements, suggesting that explanation strategies for this group can be more layered or adaptive in depth. However, intermediates still showed overload when exposed to high-density explanations, reinforcing the need for scalable complexity.

Quantitative validation through Kruskal-Wallis and Dunn's tests confirmed that these perceptions were not just subjective impressions. Statistically significant differences emerged in metrics like trust, skill-level alignment, and visual overload. This provided strong empirical backing for the tiered explanation design and the differentiation between user groups.

Despite these successes, some limitations emerged. While the visual elements were broadly rated as helpful, they occasionally contributed to a sense of clutter, especially among beginners. This points to the need for more adaptive visual complexity, where simpler overlays could be used for novices and richer visuals for experts. Additionally, although expertise level was pre-assigned in this study, real-world systems would benefit from mechanisms to infer the expertise level of the user. Lastly, the greatest limitation is the size of the user study. Although the results revealed clear trends, and were supported by statistically significant differences in many cases, the limited number of users restricts the generalizability of the findings. A larger more diverse sample would provide greater confidence of the findings, and enable further analyses (e.g, by role, age, familiarity with technology).

7 Responsible Research

7.1 Reproducibility and Transparency

To ensure reproducibility, all core code components, have been implemented in Python in a modular manner. Documentation for each component is provided via comments and a README file describing the architecture and usage. All source code and evaluation materials will be made publicly available in a dedicated GitHub[1] repository.

Furthermore, the pose estimation model and explanation modules were built using widely adopted, open-source libraries such as MediaPipe, OpenCV, and SHAP. These choices were made to ensure accessibility and reproducibility without reliance on proprietary software.

User survey data was collected through an online form, and stored in an anonymized formation. While the video data used in the prototype includes non-personal cricket footage, any data sourced externally was appropriately credited and used in compliance with publicly available licenses.

7.2 Ethical Considerations

The project involved a human-centered user study. To ensure ethical compliance, informed consent was obtained from all participants before collecting data. Participants were informed about the purpose of the study, and how their data would be used. No personally identifiable information was collected, and all data was processed anonymously.

This research focuses exclusively on domain-specific expertise, ensuring that the explanation models cannot be used for profiling or discriminatory purposes. Additionally, the system does not attempt to infer user expertise levels automatically, suppressing any ethical concerns regarding profiling or misclassifications. The user's expertise level is taken as given based on self-reporting, which avoids problematic inference mechanisms.

No deceptive practices or manipulation were involved in presenting explanation models. Each participant viewed all explanations aligned with all expertise level and provided feedback voluntarily. The premise of the project was not mentioned in the form, to avoid any biases towards explanations.

7.3 Societal and Academic Impact

This project contributes to making AI systems in cricket more accessible and interpretable. In particular, it addresses the often overlooked requirement for domain-specific, expertise-sensitive explanation systems. The personalized explanation framework developed here could have applications beyond cricket, extending to other sports, as well as other settings where expertise-sensitive explanation systems would be useful, such as training settings and education.

[1] https://github.com/AnshSharmaKumar01/generating_cricket_shot_explanations

From an academic perspective, this research reinforces the value of user-centered XAI design and adds to the literature by demonstrating its feasibility in the pose estimation domain. By integrating human factors and domain knowledge, it bridges a crucial gap between AI development and real-world applicability.

7.4 Limitations

Despite the care taken in designing and executing this study, several limitations must be acknowledged.

First, the sample size of 17 participants is relatively small, which limits the statistical power of the evaluation and may affect the generalizability of the results. This limitation was partly mitigated by using non-parametric statistical methods that are more robust to small sample sizes. Future work should validate these findings using large, more diverse groups.

Second, the expertise level of the users was based on self-reporting, which introduces subjectivity and potential misclassifications. However, this decision was motivated through practicality and ethics. Implementing a system that infers the users' expertise level was out of the scope of this research, furthermore self-reporting avoids intrusive or automated classification.

Third, the explanations were only tested on a limited set of cricket actions. While this was sufficient for a proof-of-concept, broader generalization would require evaluating the prototypes across a wider range of techniques and body types.

Lastly, although SHAP was used for generating expert-level explanations, this study did not perform deep model interpretability analysis or compare across multiple XAI models. The chosen tool was selected for accessibility and interpretability, but future research could explore complementary or alternative explanation mechanisms for improved personalization.

8 Conclusions and Future Work

This research explored how explanations generated by AI pose estimation systems can be tailored to different levels of cricket expertise. The goal was to determine whether literature regarding explanation generation could be applied to cricket, furthermore we investigate whether explanation formats adapted to user expertise could improve user experience and perceived usefulness in a sports training context. Through a combination of literature reviews, taxonomy development, prototype implementation, and user evaluation, the project confirms that expertise-level tailoring is not only beneficial but arguably essential for designing explainable systems in cricket analytics, confirming that existing literature can be applied to cricket.

The study found that users consistently rated explanation formats designed for their own expertise level as more useful, appropriate, and easier to understand. Expert users favored detailed, data-driven formats, whule beginners preferred intuitive visual overlays and high-level conceptual framing. Intermediate users sat between these extremes, benefiting from layered or moderately detailed explanations. These results were validated both qualitatively and statistically, with Kruskal-Wallis and Dunn's tests showing significant inter-group differences. These findings affirm the value of personalized XAI approaches in sports domains, extending existing work to the underexplored field of cricket pose estimation.

However, the project had several limitations. Most notably, the user study involved only 17 participants. While the results were compelling, the small size limits statistical power and generalizability. Furthermore, prototypes were developed in advanced, using existing footage, rather than real time generation, future systems could allow users to generate feedback live based on their own footage, perhaps also allowing the expertise of the user to be inferred by the system, rather than provided by the user.

Future work should aim to scale both the implementation and the evaluation. Expanding the participant pool would allow for more insights, such as analyzing the impact of experience within a giem expertise level (eg. beginner with prior coaching vs. true novice). Incorporating progressive disclosure or interactive elements in explanations could also improve engagement across user levels. Additionally, integrating adaptive systems that can assess user performance and adjust feedback dynamically would brig this work closer to applications in real-world coaching.

Overall, this study demonstrates the feasibility and necessity of expertise-sensitive XAI design in cricket and offers a blueprint for how such systems can be evaluated and iteratively improved. The insights gained through this study can be used as a foundation for future research in explainable sports analytics and personalized AI training tools.

References

[1] Ömer Konak Arıcı and Ahmet J. Güneş. Accuracy detection in some sports training using mediapipe pose estimation. *Bitlis Eren University Journal of Science and Technology*, 13(2):133–158, 2023. URL https://dergipark.org.tr/en/download/article-file/3277324.

[2] Valentin Bazarevsky, Ivan Grishchenko, Karthik Raveendran, Tyler Zhu, Fan Zhang, and Matthias Grundmann. Blazepose: On-device real-time body pose tracking, 2020. URL https://arxiv.org/abs/2006.10204.

[3] Saša Brdnik, Vili Podgorelec, and Boštjan Šumak. Assessing perceived trust and satisfaction with multiple explanation techniques in xai-enhanced learning analytics. *Electronics*, 12(12), 2023. ISSN 2079-9292. doi: 10.3390/electronics12122594. URL https://www.mdpi.com/2079-9292/12/12/2594.

[4] Michelene T.H. Chi, Paul J. Feltovich, and Robert Glaser. Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2):121–152, 1981.

[5] Cristina Conati, Oswald Barral, Vanessa Putnam, and Lea Rieger. Toward personalized xai: A case study in intelligent tutoring systems. *Artificial Intelligence*, 298:103503, 03 2021. doi: 10.1016/j.artint.2021.103503.

[6] Creative Club Productions. India vs australia 2023 - cricket highlights. https://www.youtube.com/watch?v=Xy1ntclhkZM, 2023. Accessed: 2025-06-21.

[7] Bhat Dittakavi, Vinay P. Namboodiri, and C.V. Jawahar. Pose tutor: An explainable system for pose correction. In *CVPR Workshops*, 2022. URL https://openaccess.thecvf.com/content/CVPR2022W/XAI/html/Dittakavi_Pose_Tutor_An_Explainable_System_for_Pose_Correction_CVPRW_2022_paper.html.

[8] Robert R. Hoffman, Shane T. Mueller, Gary Klein, and Jordan Litman. Measures for explainable ai: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-ai performance. *Frontiers in Computer Science*, Volume 5 - 2023, 2023. ISSN 2624-9898. doi: 10.3389/fcomp.2023.1096257. URL https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2023.1096257.

[9] Slava Kalyuga, Paul Chandler, and John Sweller. Levels of expertise and instructional design. *Human Factors*, 40(1):1–17, 1998.

[10] Bruno Martinović. Explaining cricket shot techniques using pose estimation with explainable ai. https://github.com/bruno-martinovic/XAI-in-cricket-shot-classification.git, 2025. Bachelor's Thesis, TU Delft, 2025.

[11] Vimla Patel and Gary J Groen. Cognitive mechanisms in expert clinical reasoning: a literature review. *Instructional Science*, 28(1):63–87, 2000.

[12] Andria Procopiou and Andriani Piki. The 12th player: Explainable artificial intelligence (xai) in football: Conceptualisation, applications, challenges and future directions. pages 213–220, 01 2023. doi: 10.5220/0012233800003587.

[13] Yufei Qiu, Zhaoyu Han, Nikhil Goyal, Jian Wu, and Krishna Narasimhan. Xpose: explainable human pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[14] Roboflow. MediaPipe Pose Estimation Model. https://roboflow.com/model/mediapipe, 2025. Accessed: June 21, 2025.

[15] Yao Rong, Peizhu Qian, Vaibhav Unhelkar, and Enkelejda Kasneci. I-cee: tailoring explanations of image classification models to user expertise. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'24/IAAI'24/EAAI'24. AAAI Press, 2024. ISBN 978-1-57735-887-9. doi: 10.1609/aaai.v38i19.30152. URL https://doi.org/10.1609/aaai.v38i19.30152.

[16] Kruti Shah, Manasvi Gupta, Nilesh Patil, Sridhar Iyer, Fatema Dolaria, and Rishita Koshtiz. XAI-driven yoga pose analysis and correction in real time. *Educational Administration: Theory and Practice*, 30(5): 14735–14741, 2024.

[17] Hafeez Ur Rehman Siddiqui, Faizan Younas, Furqan Rustam, Emmanuel S. Flores, Julién B. 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):6839, 2023. doi: 10.3390/s23156839.

[18] Agamdeep Singh, Sujit PB, and Mayank Vatsa. Poze: Sports technique feedback under data constraints, 11 2024.

[19] Tommaso Turchi, Alessio Malizia, Fabio Paternò, Simone Borsci, Alan Chamberlain, and Andrew Fish. Adaptive xai: Advancing intelligent interfaces for tailored ai explanations (2nd edition). In *Companion Proceedings of the 30th International Conference on Intelligent User Interfaces*, IUI '25 Companion, page 172–174, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400714092. doi: 10.1145/3708557.3716159. URL `https://doi.org/10.1145/3708557.3716159`.

[20] Niels Van Berkel, Artur Rapp, Judy Kay, and Glenn A. Wiggins. Human-centered explainable ai: Progress and challenges. *ACM Transactions on Interactive Intelligent Systems*, 11(3-4):1–36, 2021.

A Explanation Taxonomy

Below is the table explaining what model exposure means in this context, with an example of how the model exposure varies with the different expertise levels.

Table 3: Model logic exposure by expertise level

User Level	Model Logic Exposure	Example Explanation
Beginner	Hidden	"Bend your front knee more" – no mention of *why* or how the model chose this.
Intermediate	Moderate	"Front knee = 40°, ideal = 55°; flagged due to significant deviation" – includes some logic about thresholds.
Expert	High	"Model output based on 35% contribution from front knee angle. Shapley value ranking: knee > bat lift > elbow." – shows internal importance weights or decision paths.

B Prototypes

B.1 Relevant Keypoints

Table 4 shows the angles that were computed, and the keypoints that were used to compute those angles. These keypoints are the ones that are displayed in the visual overlay.

Joint Angle	Keypoints Used
Front Knee Angle	left_hip, left_knee, left_ankle
Back Knee Angle	right_hip, right_knee, right_ankle
Front Elbow Angle	left_shoulder, left_elbow, left_wrist
Back Elbow Angle	right_shoulder, right_elbow, right_wrist
Shoulder Rotation	left_shoulder, right_shoulder
Hip Rotation	left_hip, right_hip
Spine Forward Lean	left_hip, right_hip, left_shoulder, right_shoulder
Head Tilt	left_ear, right_ear
Bat Angle	right_wrist, right_thumb

Table 4: Mapping between computed joint angles and the keypoints used for their calculation.

B.2 Computing Relevant Angles

Since the joints were not used, but rather the angles between major joints, this is how those angles were computes:

Computation of Joint Angles

This section details how each joint angle is computed from pose keypoints. The angles are derived using the textttcalculate_angle function or basic trigonometric relationships involving the relevant joint coordinates. The coordinates are assumed to be 2D points of the form (x, y).

1. Knee Angles

The knee angle is computed using three keypoints: hip, knee, and ankle.

$$\text{Knee Angle} = \angle(\text{Hip}, \text{Knee}, \text{Ankle})$$

Front Knee: `LEFT_HIP`, `LEFT_KNEE`, `LEFT_ANKLE`
Back Knee: `RIGHT_HIP`, `RIGHT_KNEE`, `RIGHT_ANKLE`

2. Elbow Angles

Computed similarly using shoulder, elbow, and wrist keypoints.

$$\text{Elbow Angle} = \angle(\text{Shoulder}, \text{Elbow}, \text{Wrist})$$

Front Elbow: `LEFT_SHOULDER`, `LEFT_ELBOW`, `LEFT_WRIST`
Back Elbow: `RIGHT_SHOULDER`, `RIGHT_ELBOW`, `RIGHT_WRIST`

3. Shoulder Rotation

Measured as the angle between the shoulder line and the horizontal axis:

$$\theta_{shoulder} = \left| \tan^{-1}\left(\frac{y_r - y_l}{x_r - x_l}\right) \right|$$

Where: (x_l, y_l) and (x_r, y_r) are the coordinates of `LEFT_SHOULDER` and `RIGHT_SHOULDER`.

4. Hip Rotation

Identical to shoulder rotation, but using hip keypoints:

$$\theta_{hip} = \left| \tan^{-1}\left(\frac{y_r - y_l}{x_r - x_l}\right) \right|$$

Where: (x_l, y_l) and (x_r, y_r) are the coordinates of `LEFT_HIP` and `RIGHT_HIP`.

5. Spine Forward Lean

Angle of the vertical line connecting the midpoint of hips and shoulders with respect to the vertical axis:

$$\theta_{spine} = \left| \tan^{-1}\left(\frac{x_s - x_h}{y_s - y_h}\right) \right|$$

Where: (x_s, y_s) = midpoint of shoulders, (x_h, y_h) = midpoint of hips.

6. Head Tilt

Angle between the ears, interpreted as the tilt of the head:

$$\theta_{head} = \left| \tan^{-1}\left(\frac{y_r - y_l}{x_r - x_l}\right) \right|$$

Where: (x_l, y_l) and (x_r, y_r) are coordinates of `LEFT_EAR` and `RIGHT_EAR`. If not detected, shoulder points are used.

7. Bat Angle

Estimated from the line between the right wrist and right thumb:

$$\theta_{bat} = \left| \tan^{-1}\left(\frac{y_{thumb} - y_{wrist}}{x_{thumb} - x_{wrist}}\right) \right|$$

Where: (x_{wrist}, y_{wrist}) and (x_{thumb}, y_{thumb}) are the coordinates of `RIGHT_WRIST` and `RIGHT_THUMB`.

Key Definitions

- $\angle(A, B, C)$ denotes the internal angle at point B formed by points A, B, and C.
- $\tan^{-1}$ is the arctangent function, giving angle in degrees.
- All angles are computed in degrees and made absolute to ignore directionality.

B.3 Feedback Templates

The feedback templates were used to produce the textual feedback. There were three different templates, one for each expertise level, as their explanation needs are different.

B.3.1 Beginner Feedback Template Example

The beginner feedback was simple one or two sentences of feedback. The feedback was split into 2 categories on whether the angles needed to increase or decrease, then further split into 3 more categories depending on how much the angles have to change. This way the feedback had more flexibility for the actual feedback, depending on their specific form. Table 5 shows a sample of the feedback for beginner level prototypes. The rest of the feedback template can be found in the github.

Table 5: Feedback for `front_knee_angle` during a cover drive

Shot Type	Joint Angle	Condition	Severity	Feedback
cover_drive	front_knee_angle	too_low	small	Bend your front knee just a bit more for stability.
cover_drive	front_knee_angle	too_low	medium	Bend your front knee further to improve your shot balance.
cover_drive	front_knee_angle	too_low	large	You're almost standing upright-bend your front knee much more to stabilize your body.
cover_drive	front_knee_angle	too_high	small	Slightly straighten your front knee for better weight transfer.
cover_drive	front_knee_angle	too_high	medium	Your front knee is overly bent-straighten it to maintain control.
cover_drive	front_knee_angle	too_high	large	Your front leg is too collapsed-keep it more extended to avoid falling over.

B.3.2 Intermediate Feedback Template Example

The intermediate feedback template provided a more quantitative and personalized explanation to users. Each feedback line included both the user's current joint angle and the ideal angle, allowing users to understand exactly how their posture deviated. This format relied on a dictionary-style structure, mapping each joint angle and shot type combination to custom feedback statements. The template dynamically filled in {user} and {ideal} values using runtime pose estimation data. This version gave users a clearer and more instructive path to improvement while still maintaining short, interpretable sentences.

Table 6: Intermediate feedback for `front_knee_angle` during a cover drive

Condition	Template
too_low	Your front knee is at {user} degrees, but it should be around {ideal} degrees. Try bending it more to get a stable base.
too_high	Your front knee is at {user} degrees, ideal is {ideal} degrees. Straighten it a bit to avoid collapsing forward.

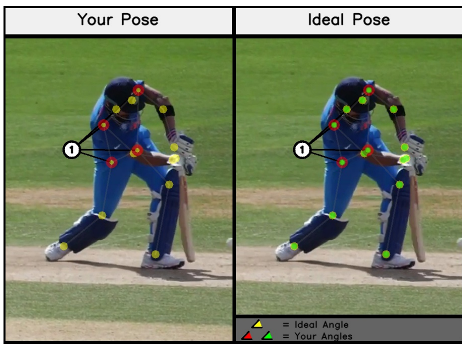
B.3.3 Expert Feedback Template Example

Since the focus of the expert level prototypes was not the text, but rather the data heavy format, the feedback was kept simple and short, so that the data would be the main focus.

Table 7: Expert feedback for `front_knee_angle` during a cover drive

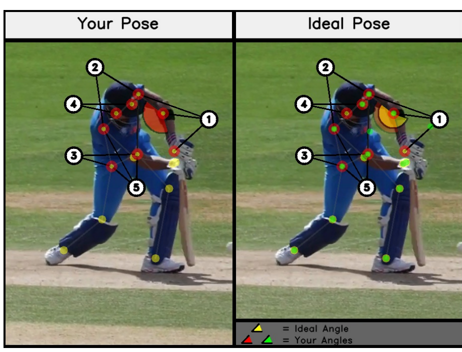
Condition	Template
too_low	Bend your front knee deeper to lower your center of gravity and stabilize your stance.
too_high	Straighten your front knee slightly to improve weight transfer and avoid over-collapsing.

B.4 Explanation Prototypes



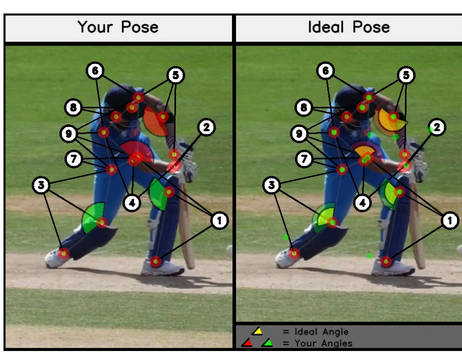
Joint	Your Angle	Ideal Angle	Deviation	SHAP (%)	Suggestion
Spine Forward Lean	176	20	155.85	9	Reduce your forward lean to maintain posture and balance.

(a) Expert – 1-point feedback



Joint	Your Angle	Ideal Angle	Deviation	SHAP (%)	Suggestion
Spine Forward Lean	176	20	155.85	9	Reduce your forward lean to maintain posture and balance.
Head Tilt	150	5	144.86	−6	Level your head to stabilize your gaze and body alignment.
Hip Rotation	155	25	129.8	18	Ease your hip turn to keep your body aligned and avoid losing power.
Shoulder Rotation	135	30	105.0	−20	Limit your shoulder rotation to maintain balance and avoid over−rotation.
Front Elbow Angle	144	90	54.14	35	Tuck your front elbow in slightly to keep the bat face square and reduce swing variance.

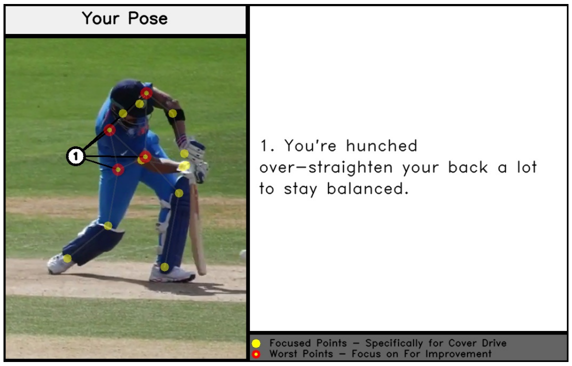
(b) Expert – 5-point feedback



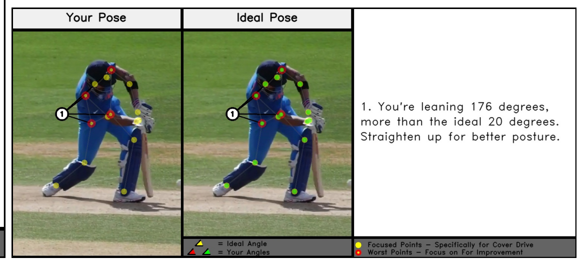
Joint	Your Angle	Ideal Angle	Deviation	SHAP (%)	Suggestion
Spine Forward Lean	176	20	155.85	9	Reduce your forward lean to maintain posture and balance.
Head Tilt	150	5	144.86	−6	Level your head to stabilize your gaze and body alignment.
Hip Rotation	155	25	129.8	18	Ease your hip turn to keep your body aligned and avoid losing power.
Shoulder Rotation	135	30	105.0	−20	Limit your shoulder rotation to maintain balance and avoid over−rotation.
Front Elbow Angle	144	90	54.14	35	Tuck your front elbow in slightly to keep the bat face square and reduce swing variance.
Back Elbow Angle	147	100	46.56	5	Straighten your back elbow slightly to optimize your swing arc and power transfer.
Back Knee Angle	139	160	−20.99	−12	Engage your back knee more by bending slightly to maintain a strong base at impact.
Bat Angle	27	45	−18.43	22	Raise your bat angle to meet the ball on the sweet spot and prevent ground−hits.
Front Knee Angle	126	140	−13.88	27	Bend your front knee deeper to lower your center of gravity and stabilize your stance.

(c) Expert – 9-point feedback

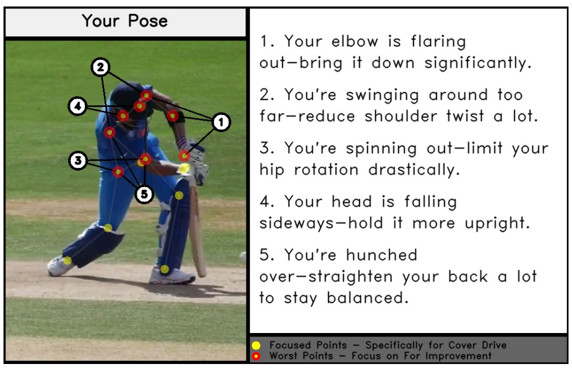
Figure 10: Expert-level prototypes across 1-point, 5-point, and 9-point focus points.



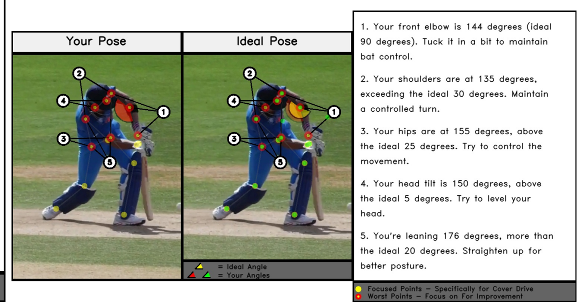
(a) Beginner – 1-point feedback



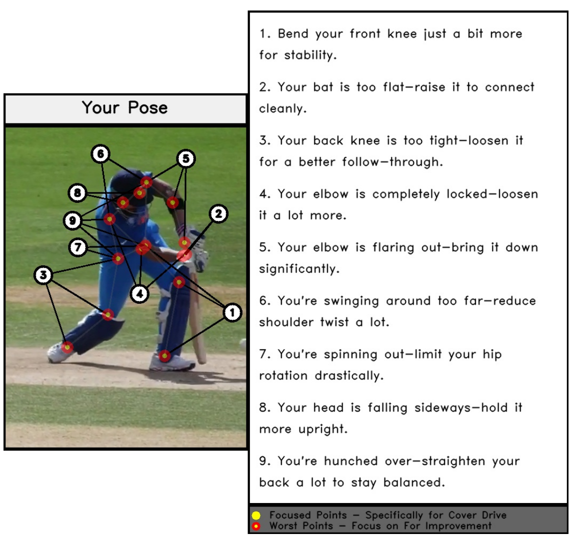
(b) Intermediate – 1-point feedback



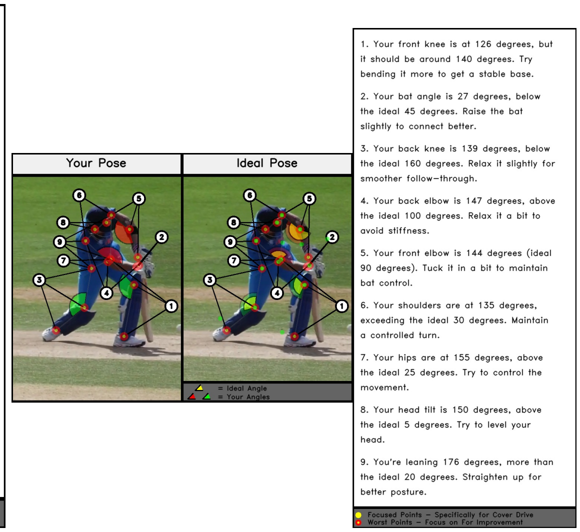
(c) Beginner – 5-point feedback



(d) Intermediate – 5-point feedback



(e) Beginner – 9-point feedback



(f) Intermediate – 9-point feedback

Figure 11: Beginner (left) and Intermediate (right) level prototypes

C Survey

The survey can be found in the following link, https://forms.gle/rj64a1tnUVk86QqZ9

D Survey Results

D.1 Background Questions Data

Consent	Expertise Level	Years Played	Play Type
Yes	Beginner	0	Viewer
Yes	Beginner	0	Viewer
Yes	Beginner	0.5	Casual Player
Yes	Beginner	2	Casual Player
Yes	Beginner	2	Casual Player
Yes	Beginner	3	Casual Player
Yes	Expert	22	Coach
Yes	Expert	13	Club Player
Yes	Expert	16	Casual Player
Yes	Expert	8	Club Player
Yes	Expert	12	Club Player
Yes	Intermediate	13	Club Player
Yes	Intermediate	5	Club Player
Yes	Intermediate	4	Casual Player
Yes	Intermediate	5	Club Player
Yes	Intermediate	6	Club Player
Yes	Intermediate	5	Club Player

Figure 12: Enter Caption

D.2 Open Questions Data

Unfortunately Open questions were only answered for the first two prototypes, as they were labeled as optional.

	Prototype 1		
Expertise Level	What are your thoughts on this explanation?	Was anything especially useful or confusing for you?	What changes should be made so that the explanation is better suited for your skill level?
Beginner	Good explanation, could be focused on a few important points that should be prioritised	When there are too many points accumulated in one spot it's confusing	Maybe points in order of what should be prioritised
Beginner			
Beginner			
Beginner			
Beginner			
Beginner			
Expert	This could be useful when training younger players, or beginners		
Expert	The explanations are too simple		
Expert	Convoluted and focused on minor details that would be automatically fixed by making one major posture change like head up and straight back	Head position being pointed out	Fewer points to focus on
Expert			
Expert	It was nice but too simple		
Intermediate	It is not bad, but could consider using more layman language to make it more user friendly.	No.	As explained above, nothing much.
Intermediate	It appears to point of flaws where there aren't any. The elbow and head position seem fine and so does the back. The only point worth noting is number 2, although that is difficult to see from a still image.	Points 1 and 5.	Some advice is better suited like play the ball under your eyes, do not let there be a gap between bat and pad. The batsman is clearly a professional and there is not much wrong with his technique.
Intermediate			
Intermediate			
Intermediate			
Intermediate			

Figure 13: Open Question Feedback for Prototype 1

	Prototype 2		
Expertise Level	What are your thoughts on this explanation?	Was anything especially useful or confusing for you?	What changes should be made so that the explanation is better suited for your skill level?
Beginner	Too many points, can be hard for beginners	The number of suggestions could be overwhelming	Less suggestions/ prioritised suggestions
Beginner			
Beginner			
Beginner			
Beginner			
Beginner			
Expert			
Expert			
Expert			
Expert			
Expert			
Intermediate			
Intermediate	There are many correct ways to play a cover drive, even among professionals, everyone has their own methods. There are certain principles that most people follow such as playing the ball late and keeping a still head but no coach tells you play a certain way, anything that feels comfortable is usually the best way.		No so specific.
Intermediate			
Intermediate			
Intermediate			
Intermediate			

Figure 14: Open Question Feedback for Prototype 2

D.3 Likert Questions Data

The next page contains the ratings of every participant for every prototype.

Prototype 1 (Beginner 5)

Expertise Level	Usefulness	Ease of Use	Trust	Appropriate Skill Level	Visual Elements Help	Too many Elements	Too many points
Beginner	Strongly Agree	Strongly Agree	Strongly Agree	Agree	Strongly Agree	Disagree	Disagree
Beginner	Agree	Agree	Agree	Agree	Neutral	Disagree	Disagree
Beginner	Strongly Agree	Strongly Agree	Agree	Strongly Agree	Agree	Disagree	Neutral
Beginner	Agree	Agree	Agree	Agree	Agree	Disagree	Neutral
Beginner	Agree	Strongly Agree	Agree	Agree	Strongly Agree	Disagree	Neutral
Beginner	Strongly Agree	Strongly Agree	Agree	Agree	Agree	Disagree	Neutral
Expert	Agree	Agree	Neutral	Neutral	Neutral	Disagree	Disagree
Expert	Neutral	Agree	Neutral	Disagree	Neutral	Strongly Disagree	Disagree
Expert	Agree	Disagree	Disagree	Disagree	Strongly Agree	Strongly Agree	Strongly Agree
Expert	Agree	Strongly Agree	Agree	Disagree	Neutral	Strongly Disagree	Strongly Disagree
Expert	Neutral	Strongly Agree	Agree	Disagree	Neutral	Strongly Disagree	Strongly Disagree
Intermediate	Agree	Strongly Agree	Agree	Strongly Agree	Strongly Agree	Neutral	Agree
Intermediate	Strongly Disagree	Neutral	Strongly Disagree	Strongly Disagree	Neutral	Strongly Agree	Agree
Intermediate	Agree	Agree	Neutral	Neutral	Agree	Disagree	Disagree
Intermediate	Agree	Agree	Neutral	Agree	Agree	Disagree	Disagree
Intermediate	Agree	Agree	Agree	Agree	Agree	Strongly Disagree	Strongly Disagree
Intermediate	Agree	Agree	Agree	Agree	Agree	Strongly Disagree	Strongly Disagree

Prototype 2 (Expert 9)

Expertise Level	Usefulness	Ease of Use	Trust	Appropriate Skill Level	Visual Elements Help	Too many Elements	Too many points
Beginner	Agree	Disagree	Agree	Agree	Agree	Agree	Agree
Beginner	Agree	Neutral	Neutral	Disagree	Neutral	Disagree	Agree
Beginner	Disagree	Strongly Disagree	Disagree	Strongly Disagree	Disagree	Strongly Agree	Strongly Agree
Beginner	Disagree	Strongly Disagree	Disagree	Disagree	Disagree	Strongly Agree	Strongly Agree
Beginner	Strongly Disagree	Disagree	Neutral	Disagree	Strongly Disagree	Strongly Agree	Strongly Agree
Beginner	Strongly Disagree	Disagree	Neutral	Disagree	Disagree	Agree	Agree
Expert	Agree	Agree	Agree	Agree	Agree	Disagree	Disagree
Expert	Agree	Neutral	Agree	Agree	Agree	Disagree	Disagree
Expert	Strongly Disagree	Strongly Disagree	Agree	Agree	Agree	Strongly Agree	Strongly Agree
Expert	Strongly Agree	Agree	Agree	Strongly Agree	Strongly Agree	Disagree	Disagree
Expert	Strongly Agree	Agree	Agree	Agree	Agree	Disagree	Disagree
Intermediate	Disagree	Neutral	Strongly Agree	Neutral	Agree	Strongly Agree	Strongly Agree
Intermediate	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Agree	Strongly Agree
Intermediate	Neutral	Disagree	Neutral	Agree	Agree	Agree	Neutral
Intermediate	Neutral	Disagree	Neutral	Neutral	Disagree	Agree	Neutral
Intermediate	Disagree	Strongly Disagree	Neutral	Disagree	Disagree	Agree	Neutral
Intermediate	Disagree	Strongly Disagree	Disagree	Disagree	Disagree	Agree	Neutral

Prototype 3 (Intermediate 5)

Expertise Level	Usefulness	Ease of Use	Trust	Appropriate Skill Level	Visual Elements Help	Too many Elements	Too many points
Beginner	Agree	Agree	Agree	Agree	Agree	Agree	Agree
Beginner	Agree	Agree	Agree	Disagree	Agree	Disagree	Disagree
Beginner	Agree	Agree	Neutral	Neutral	Neutral	Agree	Neutral
Beginner	Neutral	Neutral	Neutral	Disagree	Neutral	Agree	Neutral
Beginner	Agree	Neutral	Neutral	Disagree	Disagree	Agree	Agree
Beginner	Agree	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
Expert	Agree	Agree	Agree	Neutral	Agree	Strongly Disagree	Disagree
Expert	Agree	Agree	Agree	Agree	Neutral	Strongly Disagree	Disagree
Expert	Disagree	Strongly Disagree	Agree	Disagree	Disagree	Strongly Agree	Strongly Agree
Expert	Agree	Agree	Agree	Disagree	Neutral	Disagree	Disagree
Expert	Agree	Agree	Agree	Neutral	Neutral	Disagree	Disagree
Intermediate	Disagree	Neutral	Neutral	Agree	Neutral	Agree	Neutral
Intermediate	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Agree	Strongly Agree
Intermediate	Agree	Agree	Agree	Agree	Agree	Disagree	Disagree
Intermediate	Strongly Agree	Strongly Agree	Agree	Strongly Agree	Strongly Agree	Disagree	Disagree
Intermediate	Strongly Agree	Agree	Agree	Strongly Agree	Agree	Disagree	Disagree
Intermediate	Strongly Agree	Agree	Agree	Agree	Agree	Disagree	Disagree

Prototype 4 (Intermediate 9)

Expertise Level	Usefulness	Ease of Use	Trust	Appropriate Skill Level	Visual Elements Help	Too many Elements	Too many points
Beginner	Agree	Agree	Agree	Disagree	Agree	Agree	Agree
Beginner	Agree	Agree	Agree	Disagree	Agree	Agree	Agree
Beginner	Disagree	Strongly Disagree	Disagree	Neutral	Neutral	Agree	Strongly Agree
Beginner	Disagree	Disagree	Disagree	Disagree	Neutral	Agree	Strongly Agree
Beginner	Neutral	Disagree	Disagree	Disagree	Strongly Disagree	Agree	Strongly Agree
Beginner	Disagree	Disagree	Disagree	Disagree	Strongly Disagree	Agree	Strongly Agree
Expert	Agree	Strongly Agree	Agree	Agree	Agree	Disagree	Neutral
Expert	Agree	Strongly Agree	Agree	Agree	Agree	Disagree	Disagree
Expert	Strongly Disagree	Strongly Disagree	Disagree	Disagree	Disagree	Strongly Agree	Strongly Agree
Expert	Agree	Strongly Agree	Agree	Neutral	Neutral	Neutral	Neutral
Expert	Agree	Agree	Agree	Neutral	Neutral	Neutral	Neutral
Intermediate	Disagree	Neutral	Disagree	Agree	Neutral	Agree	Neutral
Intermediate	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Agree	Strongly Agree
Intermediate	Agree	Neutral	Neutral	Agree	Agree	Agree	Agree
Intermediate	Agree	Agree	Agree	Agree	Agree	Neutral	Agree
Intermediate	Agree	Agree	Agree	Agree	Agree	Neutral	Agree
Intermediate	Agree	Agree	Agree	Agree	Agree	Neutral	Agree

Prototype 5 (Expert 1)

Expertise Level	Usefulness	Ease of Use	Trust	Appropriate Skill Level	Visual Elements Help	Too many Elements	Too many points
Beginner	Agree	Agree	Agree	Agree	Agree	Strongly Disagree	Strongly Disagree
Beginner	Agree	Agree	Neutral	Agree	Agree	Strongly Disagree	Strongly Disagree
Beginner	Neutral	Neutral	Neutral	Neutral	Disagree	Agree	Strongly Disagree
Beginner	Neutral	Neutral	Neutral	Neutral	Disagree	Neutral	Disagree
Beginner	Neutral	Neutral	Disagree	Neutral	Neutral	Neutral	Disagree
Beginner	Neutral	Neutral	Disagree	Disagree	Agree	Neutral	Strongly Disagree
Expert	Neutral	Agree	Neutral	Agree	Agree	Strongly Disagree	Strongly Disagree
Expert	Neutral	Strongly Agree	Neutral	Agree	Agree	Strongly Disagree	Strongly Disagree
Expert	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Disagree	Strongly Disagree
Expert	Disagree	Agree	Neutral	Disagree	Neutral	Strongly Disagree	Strongly Disagree
Expert	Neutral	Strongly Agree	Neutral	Disagree	Neutral	Strongly Disagree	Strongly Disagree
Intermediate	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Disagree	Strongly Disagree
Intermediate	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree
Intermediate	Neutral	Neutral	Disagree	Neutral	Neutral	Strongly Disagree	Strongly Disagree
Intermediate	Neutral	Neutral	Disagree	Disagree	Neutral	Agree	Strongly Disagree
Intermediate	Neutral	Agree	Neutral	Neutral	Neutral	Agree	Strongly Disagree
Intermediate	Neutral	Agree	Neutral	Agree	Disagree	Agree	Strongly Disagree

Prototype 6 (Beginner 9)

Expertise Level	Usefulness	Ease of Use	Trust	Appropriate Skill Level	Visual Elements Help	Too many Elements	Too many points
Beginner	Agree	Agree	Agree	Agree	Agree	Disagree	Disagree
Beginner	Agree	Agree	Neutral	Agree	Agree	Agree	Agree
Beginner	Agree	Agree	Agree	Strongly Agree	Agree	Disagree	Strongly Agree
Beginner	Agree	Agree	Agree	Agree	Neutral	Neutral	Strongly Agree
Beginner	Agree	Agree	Agree	Agree	Neutral	Neutral	Agree
Beginner	Agree	Agree	Agree	Agree	Agree	Neutral	Strongly Agree
Expert	Neutral	Strongly Agree	Neutral	Disagree	Neutral	Disagree	Neutral
Expert	Neutral	Strongly Agree	Neutral	Disagree	Neutral	Strongly Disagree	Neutral
Expert	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Disagree	Strongly Disagree
Expert	Agree	Strongly Agree	Agree	Disagree	Agree	Disagree	Disagree
Expert	Agree	Strongly Agree	Neutral	Disagree	Agree	Disagree	Disagree
Intermediate	Strongly Agree	Strongly Agree	Strongly Agree	Agree	Strongly Agree	Strongly Agree	Agree
Intermediate	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Agree	Strongly Agree
Intermediate	Agree	Agree	Neutral	Agree	Agree	Agree	Agree
Intermediate	Strongly Agree	Agree	Agree	Agree	Agree	Neutral	Strongly Agree
Intermediate	Agree	Agree	Agree	Agree	Agree	Disagree	Neutral
Intermediate	Strongly Agree	Strongly Agree	Agree	Agree	Agree	Disagree	Neutral

Prototype 7 (Beginner 1)

Expertise Level	Usefulness	Ease of Use	Trust	Appropriate Skill Level	Visual Elements Help	Too many Elements	Too many points
Beginner	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Disagree	Agree	Strongly Disagree	Strongly Disagree
Beginner	Agree	Agree	Neutral	Disagree	Agree	Strongly Disagree	Strongly Disagree
Beginner	Strongly Agree	Strongly Agree	Agree	Agree	Agree	Strongly Disagree	Strongly Disagree
Beginner	Strongly Agree	Strongly Agree	Agree	Agree	Agree	Strongly Disagree	Strongly Disagree
Beginner	Agree	Strongly Agree	Agree	Agree	Agree	Disagree	Strongly Disagree
Beginner	Strongly Agree	Strongly Agree	Agree	Agree	Agree	Strongly Disagree	Strongly Disagree
Expert	Disagree	Strongly Agree	Neutral	Disagree	Neutral	Strongly Disagree	Strongly Disagree
Expert	Disagree	Strongly Agree	Neutral	Strongly Disagree	Neutral	Strongly Disagree	Strongly Disagree
Expert	Agree	Strongly Agree	Agree	Disagree	Agree	Strongly Disagree	Strongly Disagree
Expert	Disagree	Strongly Agree	Neutral	Disagree	Agree	Strongly Disagree	Strongly Disagree
Expert	Disagree	Strongly Agree	Disagree	Strongly Disagree	Neutral	Strongly Disagree	Strongly Disagree
Intermediate	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Disagree	Strongly Disagree
Intermediate	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree
Intermediate	Agree	Agree	Agree	Neutral	Agree	Strongly Disagree	Strongly Disagree
Intermediate	Neutral	Agree	Agree	Agree	Agree	Strongly Disagree	Strongly Disagree
Intermediate	Agree	Strongly Agree	Agree	Agree	Agree	Strongly Disagree	Strongly Disagree
Intermediate	Neutral	Strongly Agree	Agree	Agree	Agree	Strongly Disagree	Strongly Disagree

Prototype 8 (Expert 5)

Expertise Level	Usefulness	Ease of Use	Trust	Appropriate Skill Level	Visual Elements Help	Too many Elements	Too many points
Beginner	Agree	Agree	Agree	Agree	Agree	Disagree	Disagree
Beginner	Strongly Agree	Strongly Agree	Neutral	Agree	Agree	Disagree	Disagree
Beginner	Disagree	Disagree	Neutral	Neutral	Disagree	Agree	Neutral
Beginner	Disagree	Disagree	Disagree	Neutral	Disagree	Agree	Neutral
Beginner	Neutral	Disagree	Disagree	Neutral	Strongly Disagree	Agree	Neutral
Beginner	Disagree	Disagree	Disagree	Neutral	Strongly Disagree	Agree	Neutral
Expert	Agree	Agree	Agree	Agree	Agree	Disagree	Disagree
Expert	Agree	Strongly Agree	Agree	Agree	Agree	Disagree	Disagree
Expert	Disagree	Disagree	Agree	Neutral	Agree	Agree	Agree
Expert	Strongly Agree	Strongly Agree	Agree	Strongly Agree	Agree	Disagree	Disagree
Expert	Agree	Agree	Agree	Agree	Agree	Disagree	Strongly Disagree
Intermediate	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Agree	Disagree	Disagree
Intermediate	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Agree	Strongly Agree
Intermediate	Neutral	Neutral	Neutral	Neutral	Neutral	Agree	Neutral
Intermediate	Neutral	Disagree	Disagree	Neutral	Disagree	Agree	Neutral
Intermediate	Neutral	Neutral	Neutral	Neutral	Neutral	Agree	Neutral
Intermediate	Agree	Neutral	Neutral	Neutral	Neutral	Agree	Disagree

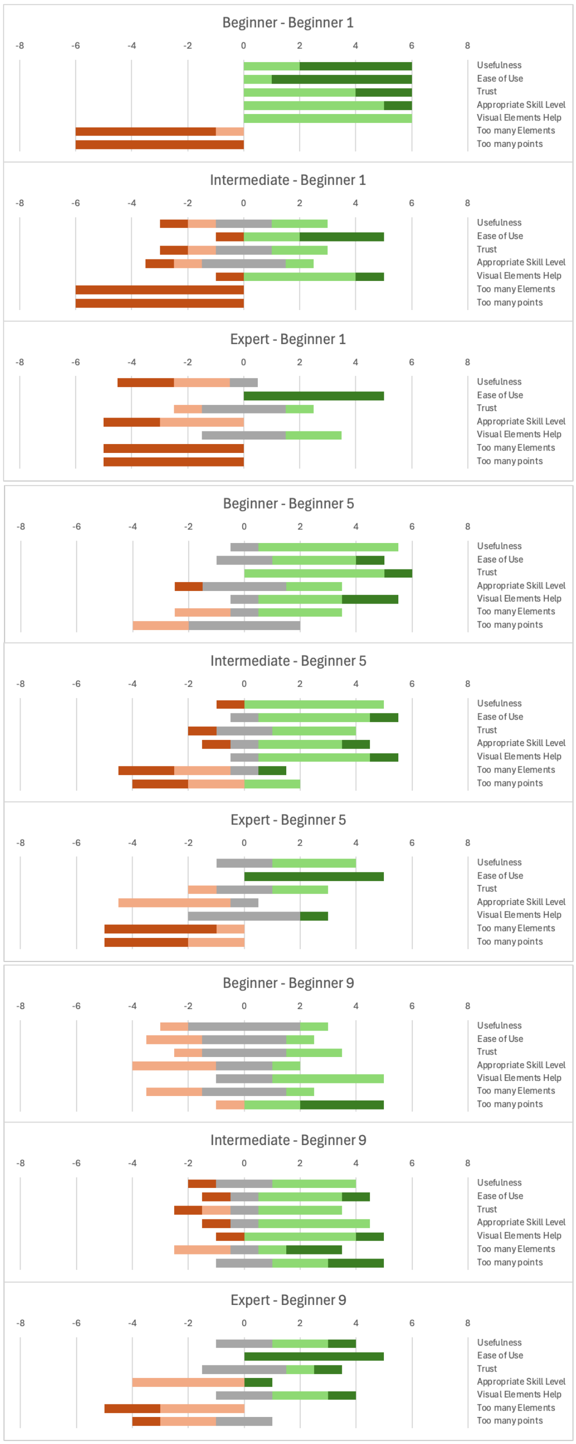
Prototype 9 (Intermediate 1)

Expertise Level	Usefulness	Ease of Use	Trust	Appropriate Skill Level	Visual Elements Help	Too many Elements	Too many points
Beginner	Agree	Agree	Agree	Agree	Disagree	Disagree	Disagree
Beginner	Neutral	Agree	Neutral	Disagree	Agree	Strongly Disagree	Strongly Disagree
Beginner	Agree	Agree	Agree	Neutral	Neutral	Neutral	Strongly Disagree
Beginner	Agree	Agree	Agree	Agree	Agree	Disagree	Disagree
Beginner	Agree	Agree	Agree	Agree	Agree	Disagree	Disagree
Beginner	Agree	Agree	Agree	Agree	Agree	Disagree	Disagree
Expert	Disagree	Strongly Agree	Neutral	Disagree	Neutral	Strongly Disagree	Strongly Disagree
Expert	Disagree	Strongly Agree	Disagree	Disagree	Neutral	Strongly Disagree	Strongly Disagree
Expert	Strongly Disagree	Disagree	Disagree	Disagree	Disagree	Strongly Disagree	Strongly Disagree
Expert	Disagree	Strongly Agree	Neutral	Disagree	Neutral	Disagree	Strongly Disagree
Expert	Disagree	Strongly Agree	Neutral	Disagree	Agree	Disagree	Strongly Disagree
Intermediate	Neutral	Neutral	Agree	Agree	Agree	Strongly Disagree	Strongly Disagree
Intermediate	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree	Strongly Disagree
Intermediate	Agree	Strongly Agree	Agree	Agree	Agree	Disagree	Strongly Disagree
Intermediate	Strongly Agree	Strongly Agree	Agree	Strongly Agree	Agree	Disagree	Strongly Disagree
Intermediate	Agree	Strongly Agree	Agree	Agree	Agree	Strongly Disagree	Strongly Disagree
Intermediate	Agree	Strongly Agree	Agree	Agree	Agree	Disagree	Strongly Disagree

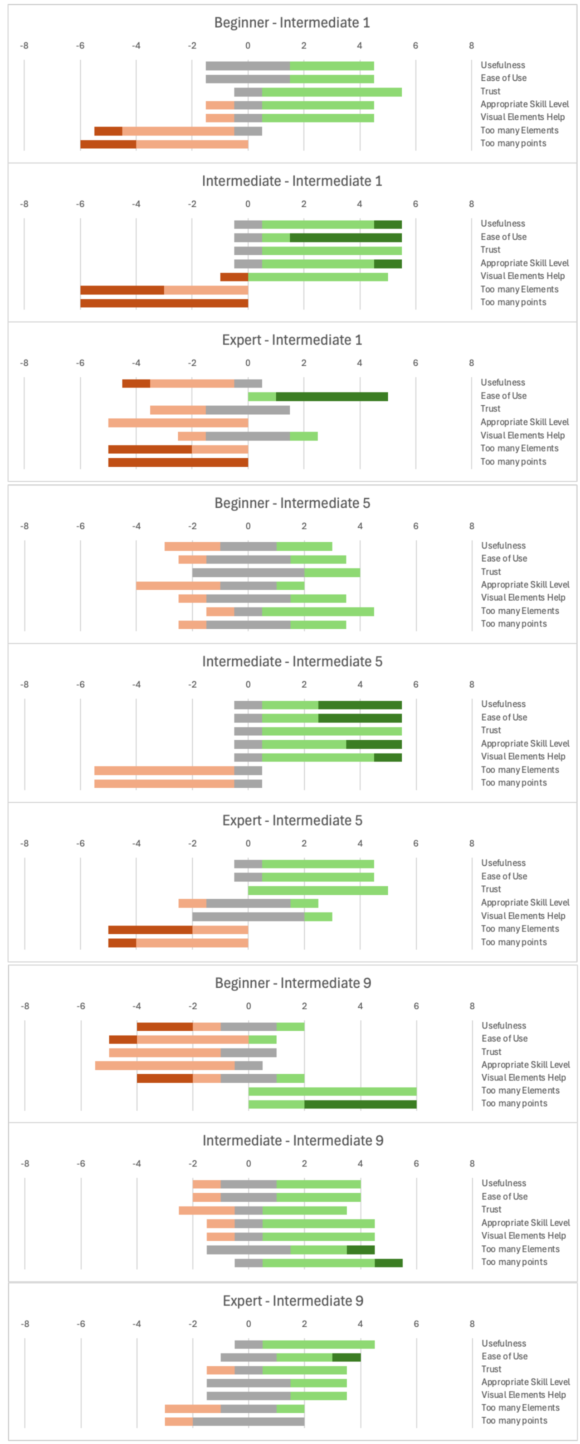
Figure 15: Likert-scale response distributions for each prototype (1–9).

D.4 Bar Charts of the Survey Results

These are the results of the Likert Scale questions from the survey. They are presented as scaled bar charts. The titles for the plots are the [User] - [Intended Prototype User + Number of Focus Points], so "Beginner - Intermediate 1" is the results of a Beginner level user, reviewing the prototype for the Intermediate level user, with 1 focus point in the prototype.

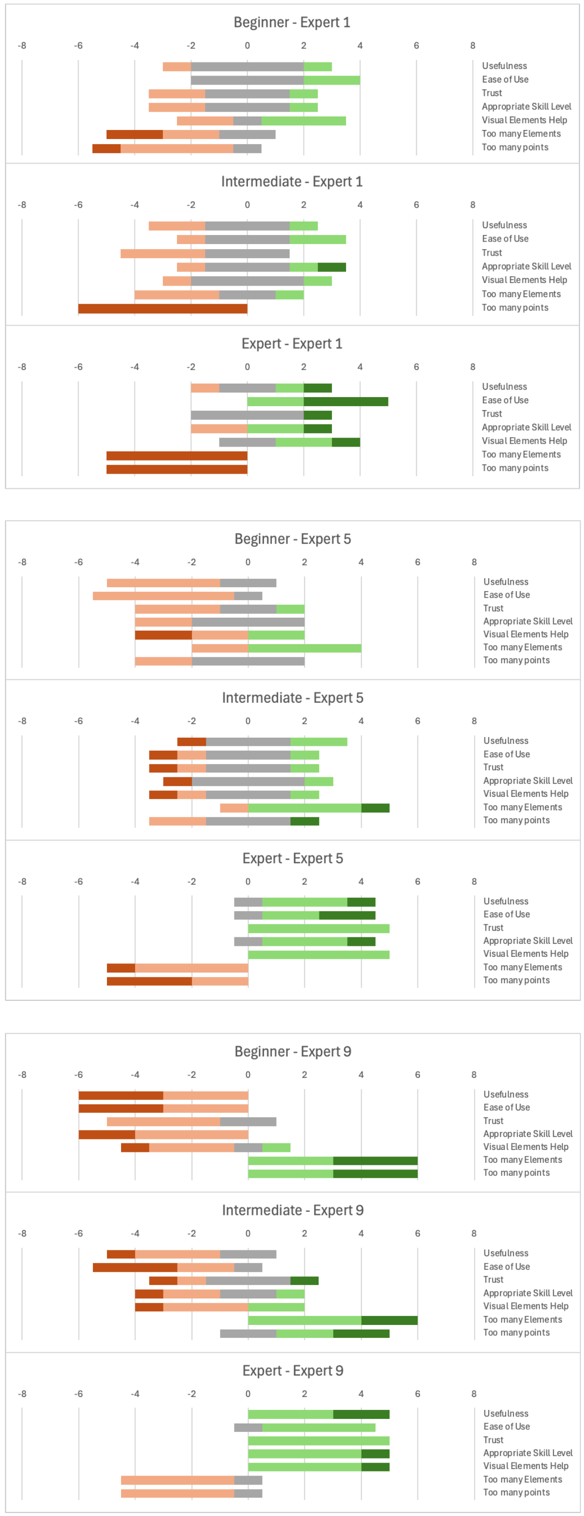


(a) Beginner level prototype ratings



(b) Intermediate level prototype ratings

Figure 16: Likert-scale results for beginner and intermediate prototypes, grouped by focus point count

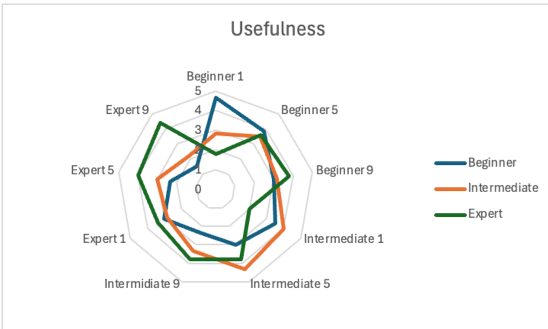


(a) Expert level prototype

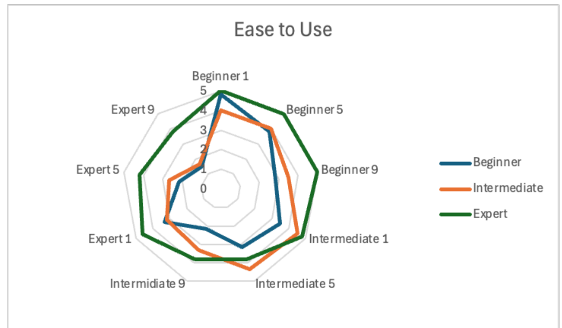
Figure 17: Likert-scale results for expert prototypes, grouped by focus point count

D.5 Radar Plots of the Survey Results

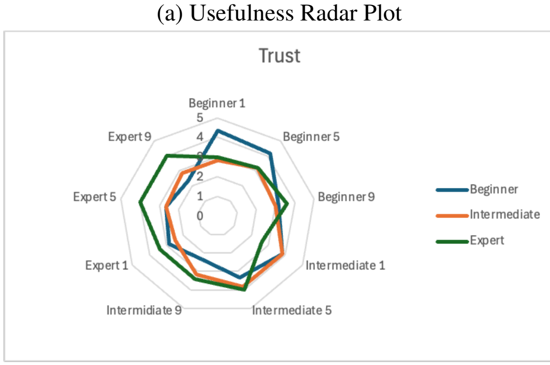
These are different visual representation of the Likert Scale questions from the survey. They plot the average ratings for each expertise level for each prototype. This gives an idea of the preferences of each expertise level, for each domain in the likery scale questions.



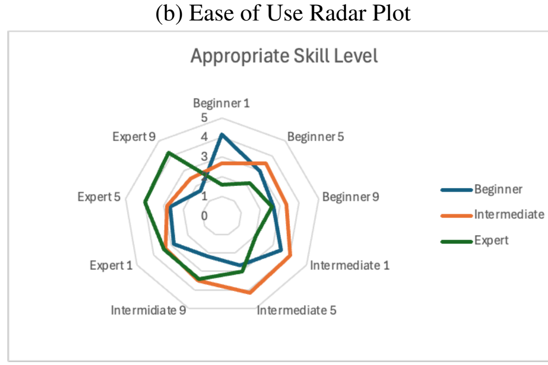
(a) Usefulness Radar Plot



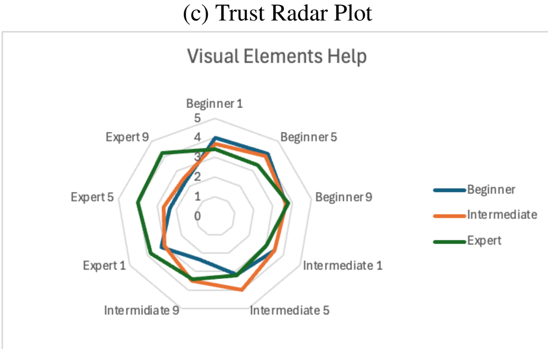
(b) Ease of Use Radar Plot



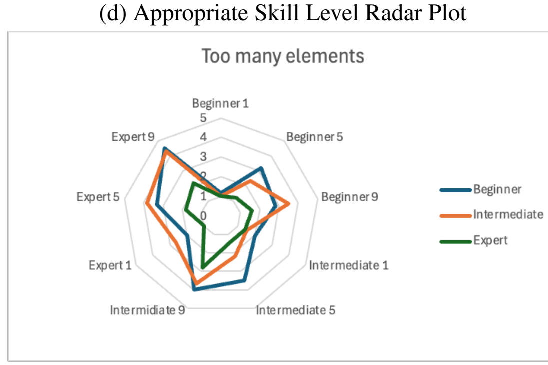
(c) Trust Radar Plot



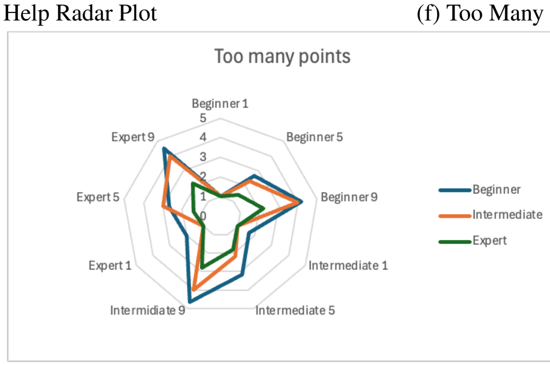
(d) Appropriate Skill Level Radar Plot



(e) Visual Elements Help Radar Plot



(f) Too Many Elements Radar Plot



(g) Too Many Points Radar Plot

Figure 18: Radar plots for all dimensions evaluated across prototypes.

D.6 Kruskal-Wallis Test and Dunn's Test p-values

Kruskal-Wallis Test p-values	Usefulness	Ease of Use	Trust	Appropriate Skill Level	Visual Elements Help	Too many Elements	Too many points
Beginner 5	0.686	0.017	0.065	0.109	0.206	0.025	0.076
Expert 9	0.003	0.005	0.016	0.004	0.026	0.004	0.005
Intermediate 5	0.047	0.057	0.046	0.024	0.077	0.004	0.009
Intermediate 9	0.066	0.039	0.114	0.013	0.133	0.041	0.003
Expert 1	0.633	0.014	0.171	0.64	0.307	0.011	0.002
Beginner 9	0.284	0.006	0.767	0.375	0.793	0.025	0.024
Beginner 1	0.004	0.139	0.017	0.003	0.141	0.4	0.4
Expert 5	0.018	0.01	0.02	0.021	0.04	0.016	0.014
Intermediate 10	0.005	0.019	0.005	0.005	0.258	0.225	0.011

Figure 19: Kruskal-Wallis Test p-scores (Green highlights significant values)

	Usefulness			Ease of Use			Trust			Appropriate Skill Level			Visual Elements Help			Too many Elements			Too many points		
Dunn Test p-values	B vs I	B vs E	I vs E	B vs I	B vs E	I vs E	B vs I	B vs E	I vs E	B vs I	B vs E	I vs E	B vs I	B vs E	I vs E	B vs I	B vs E	I vs E	B vs I	B vs E	I vs E
Beginner 5	0.877	0.41	0.499	0.758	0.008	0.019	0.053	0.038	0.82	0.443	0.175	0.037	0.737	0.09	0.169	0.212	0.007	0.128	0.371	0.024	0.159
Expert 9	0.288	0.001	0.018	0.835	0.003	0.006	0.373	0.005	0.047	0.235	0.001	0.033	0.808	0.013	0.025	0.717	0.002	0.006	0.442	0.002	0.016
Intermediate 5	0.014	0.185	0.303	0.017	0.215	0.298	0.065	0.019	0.558	0.009	0.566	0.054	0.052	0.928	0.052	0.051	0.001	0.149	0.028	0.003	0.403
Intermediate 9	0.141	0.022	0.374	0.065	0.015	0.506	0.112	0.052	0.667	0.008	0.019	0.849	0.056	0.146	0.711	0.315	0.012	0.119	0.187	0.001	0.036
Expert 1	0.684	0.571	0.34	0.808	0.014	0.007	0.398	0.284	0.06	0.421	0.413	0.959	0.627	0.294	0.13	0.251	0.057	0.003	0.002	0.004	1
Beginner 9	0.421	0.113	0.412	0.273	0.002	0.034	0.976	0.513	0.532	0.279	0.792	0.195	0.498	0.787	0.707	0.55	0.041	0.009	0.769	0.012	0.025
Beginner 1	0.023	0.001	0.276	0.165	0.556	0.056	0.013	0.016	0.977	0.031	0.001	0.187	0.831	0.067	0.103	0.234	0.256	1	0.234	0.256	1
Expert 5	0.161	0.005	0.132	0.313	0.003	0.043	0.904	0.012	0.017	0.512	0.007	0.039	0.831	0.02	0.034	0.451	0.037	0.005	0.808	0.014	0.007
Intermediate 10	0.336	0.023	0.001	0.031	0.009	0.569	1	0.004	0.004	0.403	0.018	0.002	0.723	0.214	0.114	0.175	0.112	0.768	0.008	0.012	1

Figure 20: Dunn's Test p-scores (Green highlights significant values)

Beginner vs. Expert							
Dunn Test p-values	**Usefulness**	**Ease of Use**	**Trust**	**Appropriate Skill Level**	**Visual Elements Help**	**Too many Elements**	**Too many points**
Beginner 5	0.41	0.008	0.038	0.175	0.09	0.007	0.024
Expert 9	0.001	0.003	0.005	0.001	0.013	0.002	0.002
Intermediate 5	0.185	0.215	0.019	0.566	0.928	0.001	0.003
Intermediate 9	0.022	0.015	0.052	0.019	0.146	0.012	0.001
Expert 1	0.571	0.014	0.284	0.413	0.294	0.057	0.004
Beginner 9	0.113	0.002	0.513	0.792	0.787	0.041	0.012
Beginner 1	0.001	0.556	0.016	0.001	0.067	0.256	0.256
Expert 5	0.005	0.003	0.012	0.007	0.02	0.037	0.014
Intermediate 10	0.023	0.009	0.004	0.018	0.214	0.112	0.012

(a) Dunn's Test *p*-values for Beginners vs Experts

Expert vs. Intermediate							
Dunn Test p-values	**Usefulness**	**Ease of Use**	**Trust**	**Appropriate Skill Level**	**Visual Elements Help**	**Too many Elements**	**Too many points**
Beginner 5	0.499	0.019	0.82	0.037	0.169	0.128	0.159
Expert 9	0.018	0.006	0.047	0.033	0.025	0.006	0.016
Intermediate 5	0.303	0.298	0.558	0.054	0.052	0.149	0.403
Intermediate 9	0.374	0.506	0.667	0.849	0.711	0.119	0.036
Expert 1	0.34	0.007	0.06	0.959	0.13	0.003	1
Beginner 9	0.412	0.034	0.532	0.195	0.707	0.009	0.025
Beginner 1	0.276	0.056	0.977	0.187	0.103	1	1
Expert 5	0.132	0.043	0.017	0.039	0.034	0.005	0.007
Intermediate 10	0.001	0.569	0.004	0.002	0.114	0.768	1

(b) Dunn's Test *p*-values for Experts vs Intermediates

Intermediate vs. Beginner							
Dunn Test p-values	**Usefulness**	**Ease of Use**	**Trust**	**Appropriate Skill Level**	**Visual Elements Help**	**Too many Elements**	**Too many points**
Beginner 5	0.877	0.758	0.053	0.443	0.737	0.212	0.371
Expert 9	0.288	0.835	0.373	0.235	0.808	0.717	0.442
Intermediate 5	0.014	0.017	0.065	0.009	0.052	0.051	0.028
Intermediate 9	0.141	0.065	0.112	0.008	0.056	0.315	0.187
Expert 1	0.684	0.808	0.398	0.421	0.627	0.251	0.002
Beginner 9	0.421	0.273	0.976	0.279	0.498	0.55	0.769
Beginner 1	0.023	0.165	0.013	0.031	0.831	0.234	0.234
Expert 5	0.161	0.313	0.904	0.512	0.831	0.451	0.808
Intermediate 10	0.336	0.031	1	0.403	0.723	0.175	0.008

(c) Dunn's Test *p*-values for Intermediates vs Beginners

Figure 21: Dunn's Test results split by expertise-level pair.