




Article

Explainable Graph-Based Golf Swing Analysis Integrating Club and Body Keypoints for Ball Flight Outcome Prediction

Seunghyeon Jung ^{1,†}, Minseok Kim ^{1,†}, Hyeonjin Kim ¹, Seungwon Jeong ¹, Yunseok Lee ¹, Yunji Kim ¹, Hyunse Lee ¹, Seoyoung Hong ², Gyumin Choi ², Jaerim Choi ² and Woojin Lee ^{1,*}

- ¹ Department of Computer Science and Artificial Intelligence, Dongguk University, Seoul 04620, Republic of Korea; alzkdpf23@dgu.ac.kr (S.J.); cdi6578296@dgu.ac.kr (M.K.); tkrhk8011@dgu.ac.kr (H.K.); youai058@dgu.ac.kr (S.J.); yslee0005@dgu.ac.kr (Y.L.); 2022113147@dgu.ac.kr (Y.K.); sae4394@dongguk.edu (H.L.)
- ² Kimcaddie Inc., Seoul 06752, Republic of Korea; shong6480@gmail.com (S.H.); gyumin.choi@kimcaddie.com (G.C.); jrim.choi@gmail.com (J.C.)
- * Correspondence: wj926@dgu.ac.kr
- † These authors contributed equally to this work.

Abstract

We propose a graph-based, explainable golf swing analysis framework that integrates human body keypoints with golf club keypoints to predict ball flight outcomes. We collected 321 driver swing sequences from six amateur golfers in a controlled studio setting, synchronizing monocular swing videos with TrackMan-derived ball trajectory measurements. Using a unified spatial–temporal graph that jointly models body joints and golf club keypoints, we trained graph neural networks (ST-GCN and STGAT) to perform three prediction tasks: Spin Axis and Launch Direction (classification) and Ball Speed (regression). Model performance was evaluated using AUC and accuracy for classification, and R^2 and RMSE for regression. STGAT achieved the best overall performance, reaching an AUC of 0.9188 and an accuracy of 78.33% for Spin Axis classification, an AUC of 0.7599 and an accuracy of 69.81% for Launch Direction classification, and an R^2 of 0.6925 with an RMSE of 6.4020 for Ball Speed prediction, outperforming traditional machine learning baselines. Finally, we applied Integrated Gradients to quantify the importance of both body and club keypoints across swing phases, enabling interpretable, phase-specific feedback to support individualized swing refinement.

Keywords: sports biomechanics; golf swing analysis; human body joints; graph networks; explainable AI



Academic Editor: Arkady Voloshin

Received: 27 February 2026

Revised: 4 April 2026

Accepted: 7 April 2026

Published: 14 April 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\) license](#).

1. Introduction

Human motion analysis has become an important research topic in computer vision, particularly with the advancement of skeleton-based representations that describe human movement through joint trajectories [1,2]. Recent improvements in human pose estimation (HPE) have made high-quality skeleton data widely accessible [3,4]. Graph-based models, such as Graph Convolutional Networks (GCNs), have shown strong capability in modeling structured spatial–temporal relationships among joints [5].

Golf swing performance relies on precise biomechanical coordination, where subtle variations in joint motion can significantly influence ball trajectory and flight outcomes [6]. Consequently, accurate and interpretable analysis of swing mechanics is essential for performance improvement and effective training [7].

Existing research has applied skeleton-based techniques to golf swing phase classification or comparison with expert references [8–10]. However, important limitations remain. Most approaches focus primarily on body joints and do not explicitly incorporate golf club dynamics, despite the critical role of club–body interaction in determining ball flight characteristics [11]. Moreover, prior studies rarely establish a quantitative relationship between swing kinematics and measured ball trajectory outcomes. Finally, joint-level contributions to performance are insufficiently interpreted, limiting the ability to provide actionable and individualized feedback.

To address these gaps, we propose a graph-based golf swing analysis framework that integrates body joints and club keypoints within a unified spatial–temporal representation. By jointly modeling human kinematics and equipment interaction, the proposed approach enables quantitative analysis of swing mechanics in relation to ball flight outcomes. In addition, we incorporate explainable artificial intelligence (XAI) techniques to identify phase-specific and joint-level contributions, facilitating interpretable biomechanical insights and personalized feedback.

In addition to prediction, this framework may provide a useful basis for examining how phase-specific body and club kinematics are associated with ball flight outcomes. Because golf swing performance depends on coordinated body and club motion, analyzing the influence of individual body joints and club keypoints across swing phases may help identify which components of the swing warrant closer technical evaluation. For example, phase-specific attribution patterns may help distinguish whether directional errors are more closely associated with rotational coordination and lower-body support during the backswing, impact, and finish, or whether performance differences are more closely related to club-head and upper-limb behavior around impact. By comparing good and bad swings and indicating how influential keypoints differ across phases, the proposed approach may support more structured interpretation of where corrective attention should be directed within the swing sequence.

The main contributions of this work are summarized as follows:

- We introduce a spatial–temporal graph framework that explicitly integrates golf club and body keypoints for swing modeling.
- We establish a quantitative link between swing mechanics and ball flight outcomes using synchronized motion and ball trajectory data.
- We provide interpretable, phase-specific feedback through joint-level attribution analysis, bridging data-driven modeling and practical golf training applications.

2. Related Work

2.1. Prior Vision-Based Golf Swing Analysis

Recent vision-based studies have demonstrated the feasibility of analyzing golf swings using pose- or video-based representations, particularly for swing phase classification, similarity assessment, and self-training support [8]. These studies showed that skeleton-based motion representations can provide useful information for evaluating swing form and identifying deviations from expert references.

However, several limitations remain in the recent vision-based and coaching-oriented golf swing assessment literature [10,12]. First, many existing approaches rely primarily on body-pose or joint-based representations and do not explicitly model golf club kinematics, despite the importance of coordinated club–body motion for impact efficiency and ball flight behavior. Second, much of the prior work has focused on qualitative coaching cues, similarity to expert swings, or benchmark-based discrimination between skill levels, whereas relatively fewer studies have directly modeled quantitative links between swing kinematics and objective shot outcomes such as Ball Speed, smash factor, launch conditions,

or carry distance. Consequently, the relationship between observed motion patterns and measured shot performance remains only partially characterized [6].

In this context, there is a need for a golf swing analysis framework that moves beyond pose-only assessment and explicitly models both body motion and club interaction in relation to outcome variables. Such an approach would enable a more performance-oriented interpretation of swing mechanics and provide a stronger foundation for practical feedback.

2.2. Graph-Based & Explainable Motion Analysis for Sports

Graph-based models have shown strong potential for analyzing skeleton sequences because they explicitly represent structured spatial relationships among keypoints while preserving temporal motion dynamics [9,10]. In particular, spatio-temporal graph architectures such as ST-GCN have been widely adopted for human motion understanding, and attention-based variants such as STGAT further improve the modeling of informative and non-local interactions between keypoints. These properties are especially relevant in sports motions, where coordinated movement across distant body segments can substantially influence performance [11].

In parallel, explainable artificial intelligence has become increasingly important in sports analytics, as it enables the interpretation of model predictions in biomechanically meaningful terms. Explainability is particularly valuable when the goal is not only to predict outcomes but also to identify which joints or motion phases contribute most strongly to performance-related variables [13]. Nevertheless, existing graph-based and explainable motion analysis studies have only limitedly addressed sports scenarios involving explicit body–equipment interaction and quantitatively measured performance outcomes [14,15].

Therefore, combining graph-based motion modeling with interpretable attribution analysis is a promising direction for golf swing analysis. Such a framework can better reflect the coordinated dynamics underlying swing mechanics while also supporting phase-specific interpretation linked to objective ball flight results.

3. Methods

3.1. Overall Framework

This study follows a three-stage framework for outcome-driven and explainable golf swing analysis, as summarized in Figure 1. First, swing videos are processed to extract body joint keypoints through human pose estimation (HPE) and golf club keypoints through object detection. These keypoints are integrated into a unified spatial–temporal graph representation that explicitly models both human kinematics and club–body interaction. Second, synchronized ball trajectory measurements are collected to establish a quantitative relationship between swing mechanics and ball flight outcomes. Finally, graph-based models are employed to capture spatial–temporal motion dynamics, and explainable artificial intelligence (XAI) techniques are applied to identify phase-specific and joint-level contributions, enabling interpretable and individualized feedback generation.

3.2. Data Collection

We collect swing data from six male amateur golfers in a controlled studio environment equipped with TrackMan 4 [16]. The participants span a range of technical levels, as summarized in Table 1. Golf swings can vary substantially across individuals. Accordingly, the sample represents a range of technical levels and swing postures, allowing the analysis to examine keypoint contributions associated with effective swing outcomes. Participants use two identical drivers and aim for maximum distance and accuracy toward a predetermined target. The iPhone records swing videos at 720 × 720 resolution and 30 fps, while TrackMan 4 automatically computes ball trajectory data.

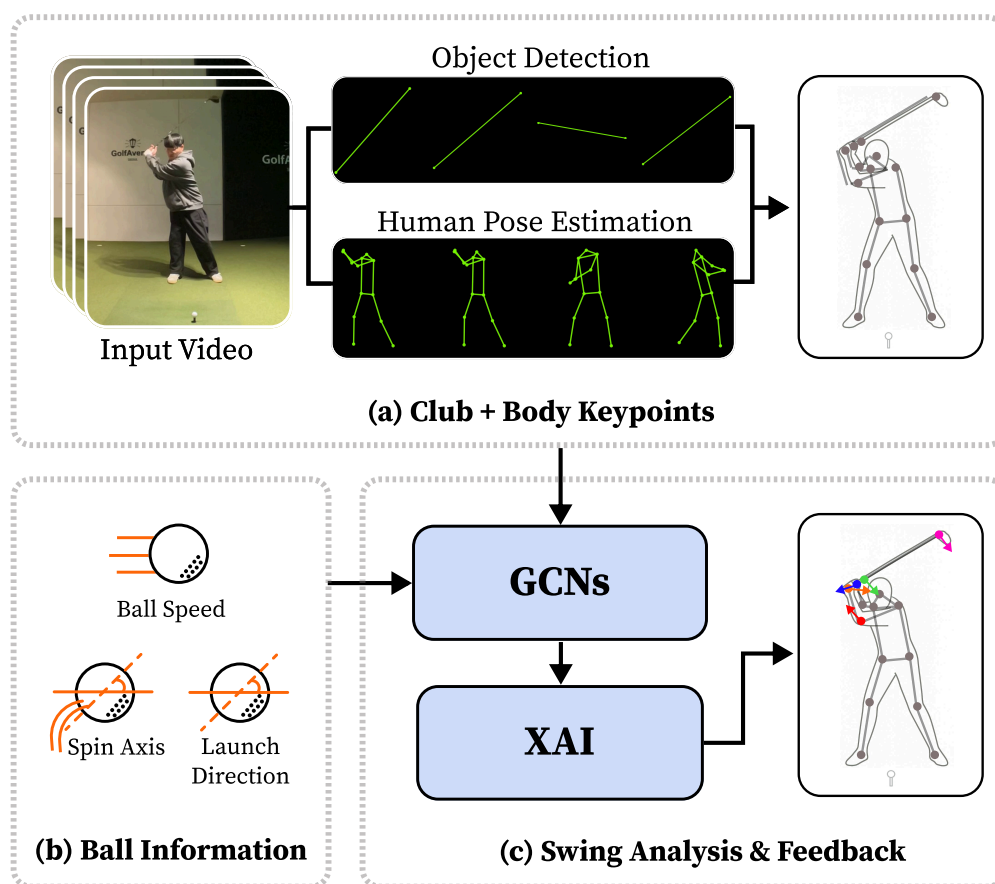


Figure 1. Overall architecture of the proposed framework. (a) Swing videos are processed through object detection and human pose estimation to extract club and body keypoints. The extracted keypoints are then organized into a graph structure. (b) Ball information is simultaneously collected. (c) The collected data are analyzed using GCNs, with XAI applied to perform swing analysis and to provide individual feedback.

A total of 563 swings were initially collected. To construct a curated dataset representing stable and analyzable driver swings, we excluded clear outliers by retaining only shots that satisfied predefined distance and directional criteria. Specifically, under-hit shots and extremely misdirected swings were removed in order to focus the analysis on representative swing mechanics rather than highly unstable impact events. After this preprocessing step, 321 swing sequences were retained for subsequent analysis.

Table 1. Detailed participant characteristics of the annotated golf swing dataset, including handicap, golf experience in years or months, and weekly training frequency. Technical level was determined by considering these self-reported characteristics together with swing outcomes.

Participant	Handicap	Golf Experience	Training Freq.	Technical Level
Player 1	15	5 years	2 times/wk	Advanced
Player 2	28	6 months	2 times/wk	Novice
Player 3	36	3 months	3 times/wk	Novice
Player 4	22	2 years	2 times/wk	Intermediate
Player 5	20	3 years	2 times/wk	Intermediate
Player 6	8	7 years	1 times/wk	Advanced

3.3. Tracking Club Information

To reflect the unique characteristics of golf swing analysis, we incorporate golf-specific factors into our study. Golf involves the use of clubs: the club-head that strikes the ball and the club-grip that connects to the golfer's body crucial keypoints. However, traditional detectors struggle to capture the high-speed club movement during the swing.

Therefore, we perform manual labeling using the Roboflow framework (<https://roboflow.com> (accessed on 3 April 2026)). We extract all frames from a video and randomly select frames from the start to the end of the swing to annotate bounding boxes for the club-head and club-grip. In total, 4812 bounding boxes for the club-head and 4639 for the club-grip are annotated. We then fine-tune a YOLO11m model on our dataset to perform object detection. The dataset is split into training, validation, and test sets in a 7:2:1 ratio to ensure robust model evaluation. To verify the reliability of the extracted club keypoints used as model inputs, we report the detection performance in Table 2.

Table 2. Evaluation results of YOLO11m on the golf swing dataset for club-head and club-grip detection. The arrows indicate that higher values correspond to better performance.

	Precision (↑)	Recall (↑)	F1-Score (↑)	mAP@50 (↑)
Overall	0.9707	0.9522	0.9613	0.9724
Grip	0.9666	0.9562	0.9614	–
Head	0.9655	0.9605	0.9630	–

3.4. Human Pose Estimation

We use a pre-trained YOLO11m-pose model to extract skeleton data from the swing video. This model extracts 17 keypoints and incorporates the club-head and club-grip coordinates. The facial keypoints consist of two eyes, two ears, and one nose, which are often occluded, leading to reduced accuracy. Thus, we only use the nose keypoint as a proxy for golfer-head movements to reduce input complexity and mitigate overfitting. Following this, Table 3 presents the definition of keypoints used in the graph structure, covering all body and club keypoints extracted from the swing video.

Table 3. Description of body joint and club keypoints used to build the graph model for analyzing golf swing mechanics.

Index	Keypoint	Index	Keypoint	Index	Keypoint
0	Golfer-Head	5	Left-Wrist	10	Right-Knee
1	Left-Shoulder	6	Right-Wrist	11	Left-Ankle
2	Right-Shoulder	7	Left-Hip	12	Right-Ankle
3	Left-Elbow	8	Right-Hip	13	Club-Head
4	Right-Elbow	9	Left-Knee	14	Club-Grip

3.5. Keypoint Graph Structure

For each swing video, we extract keypoint coordinates in the form of a tensor with dimensions (C, T, V, M) for input into the models. C denotes the x and y coordinates, along with the confidence score. T represents the number of frames, with each video is standardized to a consistent length of 368 frames. V is the number of keypoints, with 15 keypoints used for the analysis. M represents the number of individuals in the video, which is set to 1 because the data collection focuses on a single golfer.

In practice, keypoint detection is not always reliable due to factors such as occlusion and motion blur. To address this, we apply linear interpolation and the Savitzky–Golay filter to the x and y coordinates, thereby smoothing missing values and improving data quality.

3.6. Applying Integrated Gradients

This analysis applies IG to STGAT models by considering into account their graph topology. To compute IG in STGAT, unconnected keypoint pairs are masked. For these masked pairs, attention values are set to negative infinity, preventing direct gradient computation. Instead, since STGAT attention can propagate through noncontiguous keypoints via multiple paths, gradients can still be obtained indirectly.

The zero baseline commonly used in conventional IG is inappropriate for structured human movements like golf swings. Human joints can only move within specific ranges due to physical constraints, and complete zero states are anatomically impossible. Therefore, a more realistic reference is required to represent natural postures in golf swings. To address this, we divide the swing motion into eight distinct phases [17]: Address, Takeaway, Backswing, Top, Downswing, Impact, Follow-through, and Finish. We adopt the first frame of each phase as the baseline for IG calculation, setting the joint positions at the start of each phase as reference points and enabling relative measurement of changes occurring during the swing.

3.7. Data Preparation and Augmentation

All experiments are conducted under identical preprocessing, data splitting, and evaluation to ensure a fair and unbiased comparison. The dataset is split into training, validation, and test sets at a 7:2:1 ratio. To improve model generalization and address class imbalance, various data augmentation techniques are applied to both training and validation data. Specifically, Gaussian noise-based augmentation is used to enhance data diversity and mitigate the issue of limited training data. Augmentation is also applied to underrepresented classes to achieve a more balanced data distribution. These steps are taken with the goal of reducing bias and improving model performance across all classes. However, during testing, only real swing data is used to preserve the realism of the results.

3.8. Target Variables

We apply different prediction approaches depending on the target variable. The primary goal of a golf swing is to achieve both accuracy and sufficient distance. Spin Axis and Launch Direction are indicators that quantify the accuracy of the swing outcome. Specifically, Spin Axis measures how much the ball curves during flight, and Launch Direction represents the initial angle of the shot. To classify these variables, we define thresholds of -7 to 7 degrees for Spin Axis and -5 to 5 degrees for Launch Direction, based on typical shot variations such as hook, straight, and slice. Ball Speed serves as an effective measure of a golfer's ability to generate power for longer shots. Since Ball Speed is a continuous measure of shot power, we use a regression approach to predict this value.

3.9. Implementation Details

To ensure a fair comparison, all models were trained under the same conditions. The AdamW optimizer was used, with an adaptive learning rate ranging from 10^{-3} to 10^{-5} , and momentum parameters set to $\beta_1 = 0.9$ and $\beta_2 = 0.999$. A ReduceLROnPlateau scheduler was used to dynamically adjust the learning rate based on validation performance, with patience values ranging from 3 to 7 epochs depending on the experimental setup. For the loss functions, MSE loss was applied to regression tasks, while cross-entropy loss was used for classification tasks. The batch size was set to 16, and early stopping was not applied to

allow full model convergence under the same training procedure. All experiments were conducted on a single NVIDIA RTX 4090 GPU.

4. Results

4.1. Dataset Characteristics

Figure 2 illustrates the distribution of the curated swing dataset across total distance and lateral deviation for the six participants. To refine the dataset, we removed outliers and retained only swings with ball distances exceeding 137.2 m and lateral deviations within 27.4 m to either side. Following these preprocessing steps, 321 swing sequences were retained from the initial 563 recorded shots for our analysis. The retained samples are concentrated within the predefined target region, indicating that the final dataset primarily consists of representative driver swings with sufficient distance and moderate directional control. The figure also reveals inter-participant variability in shot dispersion, reflecting natural differences in swing consistency among the golfers.

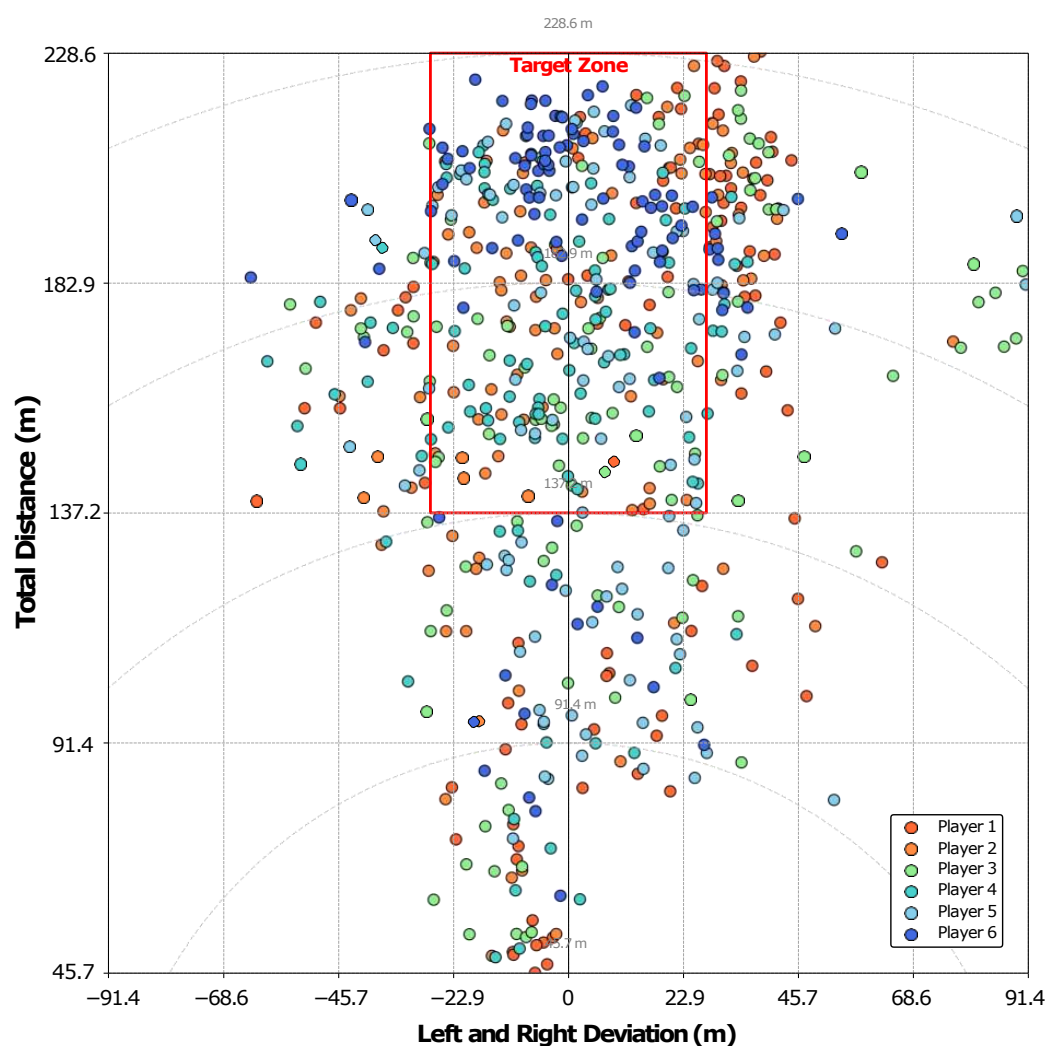


Figure 2. Distribution of the curated golf swing dataset across total distance and lateral deviation for the six participants. Each point represents one recorded shot. The red box indicates the target zone used to retain representative swings for subsequent analysis.

4.2. Overall Performance

We present the experimental results of predicting three target variables (Spin Axis, Launch Direction, and Ball Speed) using keypoints extracted from our dataset. To this end, the experiments are conducted with two graph-based models, ST-GCN and STGAT, as well as four traditional machine learning methods: Logistic Regression (LR), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), and Random Forest (RF). Table 4 compares the results across all evaluated models.

Among the evaluated models, STGAT achieves the best performance on most target variables, outperforming both traditional machine learning methods and ST-GCN. In particular, it yields the highest AUC and accuracy for Spin Axis classification and the highest accuracy for Launch Direction, while also showing competitive performance for Ball Speed regression. These results demonstrate that the graph-based spatial-temporal modeling approach is effective for representing golf swing dynamics.

Table 4. Performance comparison across models for three target variables. For classification tasks, AUC and accuracy (Acc.) are reported (higher is better). For regression, R² (higher is better) and RMSE (lower is better) are reported. Best results are highlighted in bold. The uparrow indicates that higher values correspond to better performance, while the downarrow indicates that lower values correspond to better performance.

Model	Spin Axis		Launch Direction		Ball Speed	
	AUC (↑)	Acc. (%) (↑)	AUC (↑)	Acc. (%) (↑)	R ² (↑)	RMSE (↓)
LR	0.6902	57.78	0.7200	46.67	0.6354	7.8400
XGBoost	0.6283	36.67	0.6800	36.67	0.6051	8.4696
SVM	0.6767	46.67	0.6933	43.33	0.6397	7.7939
RF	0.6933	46.67	0.7042	36.67	0.7802	6.0870
ST-GCN	0.6576	51.67	0.7767	67.92	0.5694	10.2292
STGAT	0.9188	78.33	0.7599	69.81	0.6925	6.4020

4.3. Ablation Study

We conduct experiments to investigate the impact of incorporating club information into the joint coordinate features. As shown in Table 5, adding club keypoints improves accuracy and regression performance in several settings, particularly for Ball Speed prediction and classification accuracy. However, some AUC values remain higher in the skeleton-only configuration, indicating that the effect of club information is not uniform across all evaluation metrics.

Table 5. Ablation study evaluating the effect of incorporating club keypoints. * denotes skeleton-only input without club information. Higher values indicate better performance for AUC, accuracy, and R², while lower values indicate better performance for RMSE. Best results are highlighted in bold. The uparrow indicates that higher values correspond to better performance, while the downarrow indicates that lower values correspond to better performance.

Model	Spin Axis		Launch Direction		Ball Speed	
	AUC (↑)	Acc. (%) (↑)	AUC (↑)	Acc. (%) (↑)	R ² (↑)	RMSE (↓)
ST-GCN *	0.7404	51.67	0.7910	66.04	0.5504	11.4599
ST-GCN	0.6576	51.67	0.7767	67.92	0.5694	10.2292
STGAT *	0.9437	75.00	0.8332	67.92	0.6418	6.9091
STGAT	0.9188	78.33	0.7599	69.81	0.6925	6.4020

4.4. Explainable Swing Analysis

To quantitatively evaluate the explainability of the proposed framework, we analyze the attribution patterns obtained from Integrated Gradients (IGs) and examine their relationship to swing performance outcomes. The IG analysis is conducted on the trained STGAT model. From the test dataset, we select one representative swing with a correct and confident prediction and another swing with an incorrect or low-confidence prediction. Among the 15 total keypoints, the five most influential keypoints for the best swing are identified using the absolute IG attributions. For the worst swing, the same five keypoints are evaluated to examine how their attributions differ under suboptimal mechanics. This approach enables the interpretation of how keypoints contribute to swing mechanics.

Figure 3 presents the IG analysis of our STGAT model for Spin Axis classification. In this example, the best swing is associated with five dominant keypoints: the Left-Shoulder, Left-Hip, Right-Hip, Right-Ankle, and Left-Ankle. Their attributions are concentrated mainly during the backswing, impact, and finish phases, indicating that these keypoints contribute strongly during critical portions of the swing sequence.



Figure 3. A detailed IG analysis of Spin Axis classification that compares the best and worst swings. The analysis shows how the contributions of keypoints vary across different phases of the swing. For visual clarity, the first 50 frames at the beginning of the swing and the last frames after its completion were omitted.

By comparison, the worst swing relies disproportionately on the downswing and contributes little at impact. This imbalance reveals disrupted body rotation and insufficient lower body support, leading to reduced accuracy toward the target.

To analyze the change in the IG attributions and its corresponding effect on the results, one keypoint is selected from the good swing and the trajectory for that specific keypoint is applied to replace the corresponding keypoint in the bad swing. The other 14 keypoints stay unchanged. Once the keypoint is changed, a cubic spline interpolation is applied to the time axis of the swing to reconstruct a natural trajectory. Additionally, the physical information of the club's length is maintained consistently. The analysis is repeated by changing keypoints one by one, observing and tracking the changes in both the IG and target values.

Figure 4 shows the average Ball Speed improvement and the average change in IG attribution for the top five keypoints identified in the swing analysis. The club-head exhibits the largest average improvement in Ball Speed, followed by the club-grip and upper-limb joints. Overall, the results show that modifying a small set of influential keypoints can produce measurable changes in the predicted performance outcome. These attribution patterns were further translated into phase-specific corrective feedback, which is discussed in Section 5.

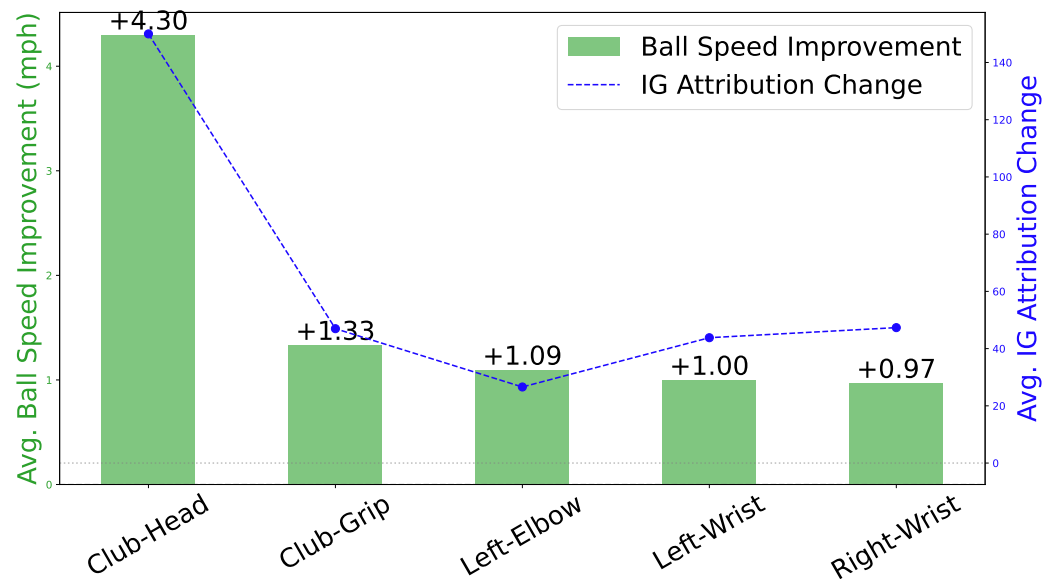


Figure 4. A visualization of the corresponding Ball Speed improvement (green) and the average change in IG attribution (blue) across the top five keypoints.

5. Discussion

5.1. Principal Findings

The present study yields three main findings. First, graph-based models were effective in capturing swing mechanics associated with ball flight outcomes. Second, incorporating club keypoints improved the physical expressiveness of the motion representation by modeling body–equipment interaction. Third, the explainability framework revealed phase-specific and joint-level attribution patterns that supported interpretable swing analysis and individualized feedback.

5.2. Effect of Incorporating Club Keypoints

While several previous golf analysis systems utilized skeletal motion features [8,9], few explicitly modeled club dynamics as independent graph nodes. As shown in Table 5, the ablation analysis supports the relevance of incorporating club keypoints within the unified graph representation. Because club-head motion directly mediates impact conditions and early ball flight characteristics [11], modeling club and body joints within the same structured framework enables a more physically grounded representation of swing mechanics.

A central implication of the ablation results is that club information should not be treated as an auxiliary cue, but as an integral component of golf swing representation. In golf, outcome variables such as Launch Direction, Spin Axis, and Ball Speed are strongly influenced by impact conditions, which are directly mediated by club-head motion and its coordination with the body. Accordingly, representing the club-head and club-grip as explicit graph nodes allows the model to capture interactions that cannot be fully inferred from body joints alone. This is particularly important because visually similar body postures may still produce different shot outcomes when the club path, face orientation, or timing of body–club coordination differs.

The use of multiple outcome variables with distinct statistical properties allows evaluation of whether the proposed representation consistently captures performance-relevant swing dynamics. Across both directional (classification) and power-related (regression) targets, integrating club–body interaction contributes to a more comprehensive modeling of outcome-relevant motion. Thus, the contribution of club modeling lies in enhancing the

physical expressiveness and outcome alignment of the swing representation rather than in isolated metric improvements.

5.3. Interpretability and Biomechanical Consistency

Compared to other explainability approaches [18,19], the use of Integrated Gradients [20] with phase-specific baselines provides a more biomechanically grounded interpretative framework. By anchoring attribution to realistic swing postures rather than zero-valued reference states, the analysis preserves biomechanical validity and mitigates the risk of generating physically implausible saliency artifacts.

In addition, the attribution results also support biomechanical interpretation of swing performance. The observed importance patterns indicate that club-related keypoints and upper-limb segments contribute substantially to performance variation, while their influence changes across swing phases. This phase dependency is important because golf performance is not determined by a single static posture, but by the sequential transfer of motion through the swing. In this sense, the proposed attribution framework provides a more meaningful interpretation than frame-level saliency alone, as it links model relevance to temporally localized mechanics that are consistent with the kinematic structure of the golf swing. This design choice enhances the credibility of the interpretability results and strengthens the overall transparency of the proposed framework.

5.4. Practical Implications

From an applied perspective, the proposed framework is relevant because it supports outcome-oriented swing assessment using relatively accessible data acquisition settings. By relying on monocular video and synchronized launch-monitor measurements [4], the method can be integrated into controlled indoor practice environments without requiring full motion-capture laboratories or wearable sensor systems [16]. This improves feasibility for routine use in coaching and performance analysis.

More importantly, the framework provides information that is directly actionable in training contexts. Coaches can use the phase-specific attribution patterns to identify not only which joints or club segments are most influential, but also when during the swing sequence corrective emphasis should be placed. For athletes, this enables a clearer distinction between outcome-relevant errors and visually noticeable but performance-neutral variations. For performance analysts, the framework offers a quantitative basis for linking swing mechanics to measurable shot results, which may support longitudinal monitoring of technical adjustments and individualized intervention strategies.

In addition to the quantitative attribution analysis, Figure 5 illustrates how the explainability results can be translated into phase-specific corrective feedback. By visualizing directional adjustment vectors for the most influential keypoints across the eight swing phases, this figure highlights the practical use of the proposed framework for individualized coaching and training. Rather than representing a direct experimental result, this visualization serves as an application-oriented interpretation of the attribution patterns identified by the model.

5.5. Limitations of the Study and Future Research Directions

This study has several limitations. The dataset is limited to six amateur golfers collected in a controlled indoor environment, and subject-disjoint evaluation on a larger and more diverse participant pool is necessary to assess generalization. In addition, the current representation relies on 2D keypoints extracted from monocular video; incorporating 3D kinematics and richer club representations may further improve biomechanical fidelity and robustness.

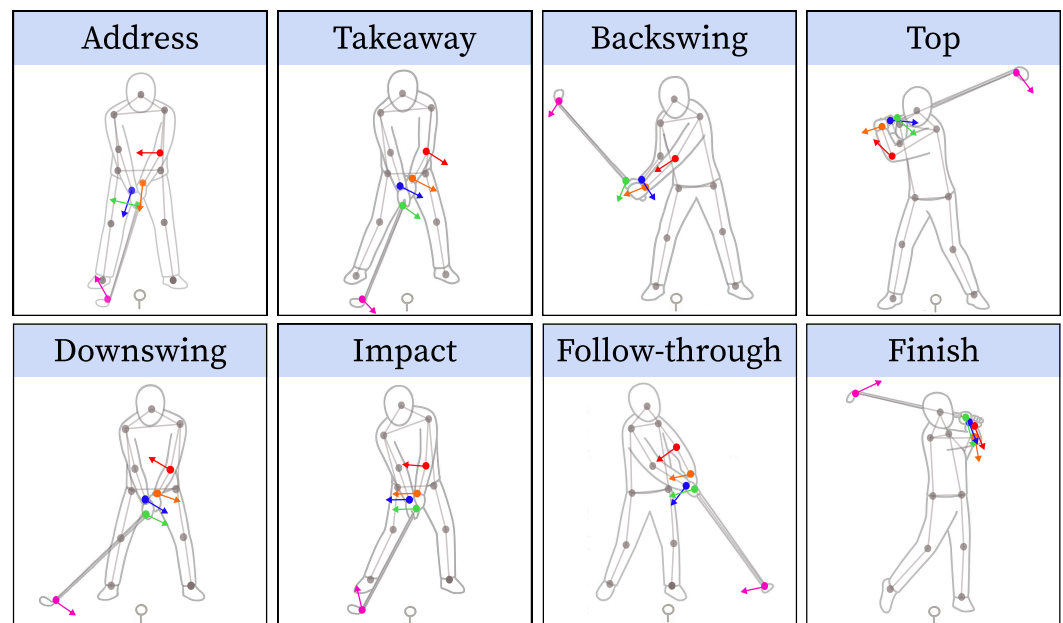


Figure 5. Keyframes illustrating the eight distinct phases of a single swing with directional vectors showing the adjustments required for the five most influential keypoints.

For Spin Axis and Launch Direction, predefined angular ranges were used to formulate directional performance as classification tasks. While this discretization enables structured evaluation, future work may explore continuous modeling or adaptive thresholding to capture finer-grained outcome dynamics. Finally, although TrackMan provides objective measurements [16], measurement variability may affect label precision. Future work will include repeated trials, uncertainty-aware modeling, and validation across multiple environments to enhance reliability and generalizability.

6. Conclusions

We propose a comprehensive framework for golf swing analysis that integrates human joint and club keypoints into a spatial–temporal graph constructed from paired swing videos and ball trajectory data. By explicitly modeling both body motion and equipment interaction, the proposed approach provides a more complete representation of swing mechanics than skeleton-only methods. The experimental results demonstrate that graph-based models effectively capture complex swing dynamics, and the ablation study further confirms that incorporating club information is essential for accurate performance evaluation.

Beyond predictive performance, a key contribution of this work lies in its interpretability. By applying explainable AI techniques, we enable detailed, joint-level biomechanical analysis across different phases of the golf swing, revealing how specific body joints and club movements contribute to swing outcomes. This interpretability allows the proposed framework to deliver individualized, phase-specific feedback, bridging the gap between data-driven modeling and practical coaching insights.

Overall, this study demonstrates that meaningful and fine-grained biomechanical analysis is achievable even for highly repetitive and fast sports motions such as the golf swing. The proposed framework not only advances golf swing analysis but also provides a generalizable methodology for motion analysis in other equipment-based sports, where understanding the interaction between the athlete and equipment is critical for performance assessment and improvement.

Author Contributions: Conceptualization, S.J. (Seunghyeon Jung), M.K. and W.L.; methodology, S.J. (Seunghyeon Jung) and M.K.; software, M.K.; validation, S.J. (Seungwon Jeong) and H.L.; formal analysis, S.J. (Seunghyeon Jung) and M.K.; investigation, Y.L. and H.L.; resources, S.H., G.C. and J.C.; data curation, S.H., G.C. and J.C.; writing—original draft preparation, S.J. (Seunghyeon Jung) and M.K.; writing—review and editing, H.K., S.J. (Seungwon Jeong) and Y.K.; visualization, H.K., Y.L. and Y.K.; supervision, W.L.; project administration, S.J. (Seunghyeon Jung); funding acquisition, J.C. and W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2025-RS-2020-II201789), and the Artificial Intelligence Convergence Innovation Human Resources Development (IITP-2025-RS-2023-00254592) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation), and the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. RS-2025-00556289).

Institutional Review Board Statement: According to the Korean Bioethics and Safety Act and its Enforcement Rule, certain categories of human-subject research may be exempt from IRB review. In particular, Article 13 of the Enforcement Rule includes, among the categories eligible for exemption: Research that uses only simple contact measurement devices or observational devices that do not involve physical changes to the human subjects.

Informed Consent Statement: Verbal informed consent was obtained from the participants. The rationale for utilizing verbal consent is that this study is a non-clinical, minimal-risk engineering study involving non-sensitive motion data, and verbal consent was considered appropriate.

Data Availability Statement: The data presented in this study are not publicly available due to privacy and ethical restrictions involving human participants. The datasets include identifiable video recordings and sensitive performance information, and their public release could compromise participant confidentiality. Therefore, the data cannot be shared publicly.

Conflicts of Interest: Authors Seoyoung Hong, Gyumin Choi, Jaerim Choi were employed by the company Kimcaddie Inc. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Kong, Y.; Fu, Y. Human action recognition and prediction: A survey. *Int. J. Comput. Vis.* **2022**, *130*, 1366–1401. [[CrossRef](#)]
2. Feng, D.; Wu, Z.; Zhang, J.; Ren, T. Multi-scale spatial temporal graph neural network for skeleton-based action recognition. *IEEE Access* **2021**, *9*, 58256–58265. [[CrossRef](#)]
3. Cao, Z.; Simon, T.; Wei, S.E.; Sheikh, Y. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2017; pp. 7291–7299.
4. Maji, D.; Nagori, S.; Mathew, M.; Poddar, D. Yolo-pose: Enhancing yolo for multi person pose estimation using object keypoint similarity loss. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2022; pp. 2637–2646.
5. Yan, S.; Xiong, Y.; Lin, D. Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*; AAAI Press: New Orleans, LA, USA, 2018; Volume 32.
6. Bourgain, M.; Rouch, P.; Rouillon, O.; Thoreux, P.; Sauret, C. Golf swing biomechanics: A systematic review and methodological recommendations for kinematics. *Sports* **2022**, *10*, 91. [[CrossRef](#)] [[PubMed](#)]
7. Jung, S.; Hong, S.; Jeong, J.; Jeong, S.; Choi, J.; Kim, H.; Lee, W. CaddieSet: A Golf Swing Dataset with Human Joint Features and Ball Information. In *Proceedings of the Computer Vision and Pattern Recognition Conference*; IEEE: Piscataway, NY, USA, 2025; pp. 5988–5996.
8. Kim, T.T.; Zohdy, M.A.; Barker, M.P. Applying pose estimation to predict amateur golf swing performance using edge processing. *IEEE Access* **2020**, *8*, 143769–143776. [[CrossRef](#)]
9. Liao, C.C.; Hwang, D.H.; Koike, H. Ai golf: Golf swing analysis tool for self-training. *IEEE Access* **2022**, *10*, 106286–106295. [[CrossRef](#)]
10. Ju, C.Y.; Kim, J.H.; Lee, D.H. GolfMate: Enhanced golf swing analysis tool through pose refinement network and explainable golf swing embedding for self-training. *Appl. Sci.* **2023**, *13*, 11227. [[CrossRef](#)]

11. Sweeney, M.; Mills, P.; Alderson, J.; Elliott, B. The influence of club-head kinematics on early ball flight characteristics in the golf drive. *Sport. Biomech.* **2013**, *12*, 247–258. [[CrossRef](#)] [[PubMed](#)]
12. Lee, M.H.; Zhang, Y.C.; Wu, K.R.; Tseng, Y.C. Golfpose: From regular posture to golf swing posture. In *Proceedings of the International Conference on Pattern Recognition*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 387–402.
13. Hu, L.; Liu, S.; Feng, W. Spatial temporal graph attention network for skeleton-based action recognition. *arXiv* **2022**, arXiv:2208.08599. [[CrossRef](#)]
14. Skublewska-Paszowska, M.; Powroznik, P.; Lukasik, E. Learning three dimensional tennis shots using graph convolutional networks. *Sensors* **2020**, *20*, 6094. [[CrossRef](#)]
15. Kim, H.; Kim, B.; Chung, D.; Yoon, J.; Ko, S.K. SoccerCPD: Formation and role change-point detection in soccer matches using spatiotemporal tracking data. In *Proceedings of the 28th Acm Sigkdd Conference on Knowledge Discovery and Data Mining*; ACM: New York, NY, USA, 2022; pp. 3146–3156.
16. Bishop, C.; Wells, J.; Ehler, A.; Turner, A.; Coughlan, D.; Sachs, N.; Murray, A. Trackman 4: Within and between-session reliability and inter-relationships of launch monitor metrics during indoor testing in high-level golfers. *J. Sport. Sci.* **2023**, *41*, 2138–2143. [[CrossRef](#)] [[PubMed](#)]
17. McNally, W.; Vats, K.; Pinto, T.; Dulhanty, C.; McPhee, J.; Wong, A. Golfdb: A video database for golf swing sequencing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*; IEEE: Piscataway, NY, USA, 2019.
18. Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why should i trust you?” Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: New York, NY, USA, 2016; pp. 1135–1144.
19. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4768–4777.
20. Sundararajan, M.; Taly, A.; Yan, Q. Axiomatic attribution for deep networks. In *Proceedings of the International Conference on Machine Learning*; JMLR: Sydney, Australia, 2017; pp. 3319–3328.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.