Gait-Based Prediction of Penalty Kick Direction in Soccer

David Freire-Obregón[a], Oliverio J. Santana[b], Javier Lorenzo-Navarro[c], Daniel Hernández-Sosa[d] and Modesto Castrillón-Santana[e]

SIANI, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain

Keywords: Penalty Kick Prediction, Soccer Analytics, Gait Analysis, LSTM, Action Anticipation, Sports Analytics.

Abstract: Understanding and predicting penalty kick outcomes is critical in performance analysis and strategic decision-making in soccer. This study investigates the potential of gait-based biometrics to classify the intended shoot zone of penalty takers using temporal gait embeddings extracted from multiple state-of-the-art gait recognition backbones. We compile a comprehensive evaluation across several models and datasets, including baseline models and other models such as GaitPart, GLN, GaitSet, and GaitGL trained on OUMVLP, CASIA-B, and GREW. A standardized LSTM-based classifier is trained to predict the shooting zone from video-level gait sequences, using consistent train-test splits to ensure fair comparisons. While performance varies across model-dataset pairs, we observe that certain combinations yield better predictive accuracy, suggesting that the gait representation and the training data influence downstream task performance to some degree. This work demonstrates the feasibility of using gait as a predictive cue in sports analytics. It offers a structured benchmark for evaluating gait embeddings in the context of penalty shoot zone prediction.

1 INTRODUCTION

Penalty shootouts represent one of professional soccer's most critical and psychologically demanding scenarios. Despite their brief duration, these isolated events often carry disproportionate weight in determining the outcome of tightly contested matches. In recent FIFA World Cups, over a quarter of knockout-stage games have been decided from the penalty mark, highlighting this phase's strategic and emotional significance. As a result, understanding the dynamics of penalty kicks has become increasingly relevant for players, coaching staff, and analysts.

Given the high pressure and game-deciding nature of penalties, the ability to anticipate the direction of a penalty shot could provide a strategic advantage to goalkeepers and analysts alike. While physical attributes, such as body orientation and approach angle, have been studied extensively, the subtler pre-shot movement patterns—specifically, the shooter's gait—remain underutilized in predictive modeling. Developing systems that can interpret a player's movement sequence before the shot and reliably predict shot direction could augment goalkeeper training and real-time match analytics.

Penalty kick prediction and related tasks have been approached by analyzing body motion, pose estimation, and temporal movement patterns, frequently supported by player tracking and activity recognition frameworks. For instance, random forests combined with context-conditioned motion models to detect and track players under dynamic conditions, enabling action inference (Liu and Carr, 2014). A comprehensive review of player tracking methods has outlined the challenges posed by occlusion and pose variability, pertinent to pre-kick movement analysis, such as gait (Manafifard et al., 2016). Interaction modeling between players and the ball has also been investigated, incorporating physical constraints to improve the interpretability of motion dynamics (Maksai et al., 2016). Recently, penalty scenarios have been addressed using Human Action Recognition (HAR) models (Freire-Obregón et al., 2025). However, unlike that work, which focuses solely on the running and kicking stages by cropping the sequence, our approach considers the entire penalty sequence without truncation at any moment. This decision is grounded in the observation that, although gait models are effective at capturing individual walking patterns, their

[a] https://orcid.org/0000-0003-2378-4277
[b] https://orcid.org/0000-0001-7511-5783
[c] https://orcid.org/0000-0002-2834-2067
[d] https://orcid.org/0000-0003-3022-7698
[e] https://orcid.org/0000-0002-8673-2725

Freire-Obregón, D., Santana, O. J., Lorenzo-Navarro, J., Hernández-Sosa, D. and Castrillón-Santana, M.
Gait-Based Prediction of Penalty Kick Direction in Soccer.
DOI: 10.5220/0013669900003988

In *Proceedings of the 13th International Conference on Sport Sciences Research and Technology Support (icSPORTS 2025)*, pages 142-149
ISBN: 978-989-758-771-9; ISSN: 2184-3201

ultimate goal is subject identification. Consequently, we aim to provide as many temporal and contextual cues as possible by preserving the full sequence. Despite these developments, the temporal modeling of fine-grained kinematic features, particularly gait during a penalty approach, has received limited attention. This gap highlights the potential for sequence-based representations to enhance predictive modeling in this context, mainly when used with recurrent neural architectures.

In this work, we propose a deep learning approach based on Long Short-Term Memory (LSTM) networks to predict the shot zone of a penalty taker using their gait sequence leading up to the shot. We construct a consistent experimental framework by leveraging several gait backbones pretrained with publicly available datasets for penalty kick scenarios. Each video is segmented into fixed-length sequences, and a neural network is trained to classify the final shot direction into one of three zones. To ensure fair comparisons, we apply a shared train/test split across all datasets and repeat experiments multiple times, saving only the best-performing model for each.

Our contributions are threefold: (1) we design a standardized and reproducible LSTM-based pipeline for predicting penalty kick direction leveraging gait pretrained backbones; (2) we benchmark this pipeline across multiple gait datasets using consistent data splits and repeated trials to account for variance in training outcomes; and (3) we conduct a qualitative error analysis of the best-performing models, revealing that most misclassifications occur between adjacent shoot zones. This indicates that gait patterns associated with neighboring shot directions are often subtly different and challenging to distinguish based on pre-kick motion alone.

2 RELATED WORK

Computer vision applications in sports have evolved to support coaching, broadcasting, and analytics through player tracking, event recognition, and motion analysis. In soccer, this has led to technologies like TRACAB and Hawk-Eye, which are used for player tracking and goal-line detection, respectively (ChyronHego, 2017; Innovations, 2017). These systems rely on calibrated multi-camera setups and computer vision pipelines for real-time data extraction. However, their focus is primarily on positional and event-level data, offering limited insight into biomechanical features such as gait.

Tracking players for tactical and performance analysis is a significant research focus. For instance, Manafifard et al. provide a survey of player tracking techniques in soccer, noting challenges such as occlusions, appearance similarity, and erratic movements (Manafifard et al., 2016). Techniques ranging from model-based detection to context-conditioned motion models have been proposed to tackle these problems (Liu and Carr, 2014). Still, these approaches focus on position rather than motion style, leaving gait-specific analysis underexplored.

Motion analysis of athletes has traditionally relied on marker-based motion capture systems, which are impractical for in-game scenarios. Markerless systems using multiple or single cameras have been explored to visualize motion trails or generate pose sequences (Figueroa et al., 2006). These visualizations are often used for coaching or broadcast enhancements but lack integration into predictive models. Furthermore, pose estimation accuracy under realistic conditions, such as during soccer penalty kicks, remains a technical challenge.

Event detection in sports has been another key area of study. Kapela et al. proposed methods for detecting goals, shots, and fouls through visual analysis and scoreboard interpretation (Kapela et al., 2014). Recent work has also examined the use of visual and temporal features for predicting shot outcomes, including the classification of ball-on-goal positions based on the kicker's shooting action (Artiles et al., 2024), as well as large-scale performance analyses in other sports domains such as running, highlighting the value of motion-based modeling across disciplines (Freire-Obregón et al., 2022). While this supports high-level game understanding, such methods do not typically incorporate pre-shot motion cues, such as approach gait, which may reveal a player's intention during set-pieces like penalty kicks.

Our work complements and extends this body of research by applying gait embeddings, commonly used for biometric identification, to predict shot zones in soccer penalty scenarios. Unlike previous approaches focused on spatial position or detected events, we explore how temporal motion patterns can serve as predictive features. This represents a novel intersection of gait recognition and sports analytics, leveraging insights from both domains.

3 METHODOLOGY

This section details the overall approach used to model and classify penalty kick directions based on visual motion cues. Our methodology is structured into three main components: formal problem definition, gait-based feature extraction, and shot direction

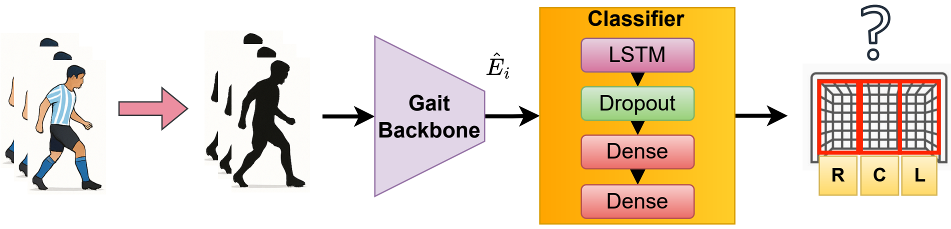

Figure 1: Pipeline overview. The kicker is isolated from each video to generate silhouette sequences encoded using a pre-trained gait model. The resulting embeddings are processed by an LSTM-based classifier to predict shot direction, R, C, and L correspond to Right, Center, and Left from the goalkeeper's perspective.

classification. By using recent advances in human motion analysis and lightweight temporal modeling, we aim to evaluate the effectiveness of gait representations in anticipating a penalty kicker's intent. Each component is described in the subsections that follow.

3.1 Problem Definition

Let there be m instances of penalty shooters, where each instance is defined as a tuple $o^{(i)} = (V^{(i)}, z^{(i)})$, for $i = 1 \dots m$. In this formulation, $V^{(i)}$ represents the input sequence derived from video data, and $z^{(i)}$ denotes the class label corresponding to the direction of the shot. The classification task involves three possible categories: $z^{(i)} \in \{\mathrm{R}, \mathrm{C}, \mathrm{L}\}$ corresponding to *Right*, *Center*, and *Left* from the goalkeeper's perspective. The objective is to learn the model parameters θ that minimize the cross-entropy loss over the dataset:

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \sum_{k=1}^{n} p(z^{(i)} = k) \log(p(\hat{z}^{(i)} = k)), \quad (1)$$

where $n = 3$ is the number of output classes, $p(z^{(i)} = k)$ is the ground truth probability for class k, and $p(\hat{z}^{(i)} = k)$ is the predicted probability for class k for instance i.

3.2 Gait Embedding

To extract motion features relevant to penalty kickers, input sequences derived from video data $V^{(i)}$ are first converted into silhouette sequences, focusing exclusively on the kicker (see Figure 1). Neither the ball nor the goalkeeper is considered, and the only visible silhouette in each sequence corresponds to the player performing the kick. Detection and tracking are performed using YOLOv8x-pose-p6 (Jocher et al., 2023) and Bot-SORT (Aharon et al., 2022), while high-precision silhouettes are obtained with SAMU-RAI (Yang et al., 2024).

Each silhouette sequence is then processed using a pre-trained gait recognition model $\mathcal{B}_{\text{GAIT}}$, trained on large-scale public datasets such as CASIA-B (Yu et al., 2006), OUMVLP (Takemura et al., 2018), and GREW (Zhu et al., 2021). Rather than producing a temporal embedding, $\mathcal{B}_{\text{GAIT}}$ encodes the sequence into a set of spatial gait descriptors corresponding to vertically partitioned body regions. This structure captures local motion characteristics across the body from top to bottom. The embedding output is defined as:

$$E_i = \mathcal{B}_{\text{GAIT}}(V^{(i)}) \in \mathbb{R}^{P \times D},$$

where P denotes the number of spatial body partitions (e.g., $P = 62$ using Horizontal Pyramid Pooling), and D is the embedding dimension per region. To improve generalization and comparability across samples, the embeddings are standardized using the training set mean μ_{train} and standard deviation σ_{train}, followed by L2 normalization:

$$\tilde{E}_i[p] = \frac{E_i[p] - \mu_{\text{train}}}{\sigma_{\text{train}}}, \quad \hat{E}_i[p] = \frac{\tilde{E}_i[p]}{\|\tilde{E}_i[p]\|_2}.$$

The resulting normalized embedding matrix $\hat{E}_i \in \mathbb{R}^{P \times D}$ is then passed to a lightweight classification model.

Although $\hat{E}_i$ is not a temporal sequence, we treat the vertical ordering of body parts (from head to foot) as a structured sequence to capture spatial dependencies. By applying a Recurrent Neural Network (RNN) such as an LSTM over the P body partitions, the model can learn hierarchical spatial interactions across regions (e.g., how lower-body motion relates to upper-body posture). This sequential processing allows the classifier to aggregate global pose information while remaining sensitive to subtle localized variations in movement style.

3.3 Shot Direction Classification

A simple temporal classification model is employed to evaluate the predictive value of gait embeddings for penalty kick direction. The objective is to learn a mapping from the normalized gait sequence $\hat{E}_i \in \mathbb{R}^{T\times D}$ to the shot label $z^{(i)} \in \{right, center, left\}$.

The model is defined as a composition of standard neural network layers:

$$f_\theta(\hat{E}_i) = \text{Softmax}\Big(W_2 \cdot \text{ReLU}\big(W_1 \cdot \text{LSTM}_{32}(\hat{E}_i) + b_1\big) + b_2\Big) \quad (2)$$

Where $\text{LSTM}_{32}(\cdot)$ denotes a unidirectional LSTM layer with 32 hidden units, $W_1 \in \mathbb{R}^{32\times 16}$ and $W_2 \in \mathbb{R}^{16\times 3}$ are weight matrices of the fully connected layers, and $b_1 \in \mathbb{R}^{16}$ and $b_2 \in \mathbb{R}^{3}$ are the corresponding bias vectors. $\text{Softmax}(\cdot)$ converts the output logits into a categorical distribution over the three classes.

This lightweight architecture is deliberately chosen to avoid overfitting and isolate the gait embeddings' representational quality. By limiting model capacity, performance differences across backbone networks can be more confidently attributed to the discriminative power of the extracted features, rather than architectural complexity. Additionally, the model's transparent structure and low computational cost make it suitable for fast benchmarking and iterative experimentation.

4 EXPERIMENTAL SETUP

Gait Backbones. Silhouette-based gait recognition methods aim to extract discriminative motion features from binary human outlines to characterize walking behavior. For this study, we adapt a range of representative architectures, including GaitBase (Fan et al., 2023), GLN Phase 1 and 2 (Hou et al., 2020), GaitGL (Lin et al., 2021), GaitPart (Fan et al., 2020), and GaitSet, for the task of predicting penalty kick direction. Though all these models process silhouette sequences, their internal mechanisms differ significantly. GaitSet treats sequences as sets of independent frames, using pooling operations to summarize features over time, but it does not explicitly encode spatial continuity. GaitPart enhances this design by focusing on horizontal body partitions through Focal Convolution, capturing localized motion patterns but introducing potential sensitivity to pose misalignment. GaitGL extends the modeling capacity by integrating global and part-based branches, alongside 3D convolutions, to extract joint spatial-temporal features. While more expressive, its increased complexity may limit performance consistency in real-world applications. The GLN models adopt a latent grouping mechanism. In a first phase (Phase 1), grouped feature representations are built, and in a second phase (Phase 2), progressive refinement layers are introduced to increase representation granularity across the network. Lastly, GaitBase provides a deeper residual network baseline demonstrating the effectiveness of capacity and depth without additional architectural innovations.

As described in Section 3, the models were evaluated using embeddings trained on three benchmark datasets: OU-MVLP, CASIA-B, and GREW. OU-MVLP provides large-scale indoor sequences under uniform conditions, while CASIA-B introduces controlled variability through clothing, carrying objects, and multi-view setups. Captured in the wild, GREW presents more realistic challenges, such as occlusion and lighting variation, which better reflect our target domain. Not all backbones were trained on every dataset, as architectural complexity and data variability required careful pairing to ensure feasible training and reliable feature extraction.

Dataset Collection and Filtering. The dataset was constructed from publicly available footage spanning international matches, professional leagues, and highlight compilations. Metadata about match level was manually checked where available. Inspired by the data acquisition strategies highlighted in prior sports vision research (Thomas et al., 2017), the collection focused on maximizing visual diversity regarding pose dynamics, camera distance, and illumination. A targeted search using terms such as "penalty-kick shootout" yielded a collection of raw video clips, each manually trimmed to retain only the relevant sequence, from the start of the run-up to the outcome of the kick (see Figure 2).

To ensure temporal consistency and viewpoint suitability, only clips recorded from optimal angles (typically side or diagonal views of the kicker) and with sufficient temporal resolution (minimum of 64 frames) were retained. This filtering process yield to a dataset to 432 penalties. Each clip was annotated with a shot direction label, where class 0 corresponds to shots aimed left, 1 to center, and 2 to right. The final label distribution was imbalanced, with 209 samples in class 0, 66 in class 1, and 157 in class 2, which reflects real-world tendencies and must be accounted for during model training and evaluation. Goalkeepers appearing in the footage were not the focus of analysis but were used for the human baseline by recording their initial dive direction.

Implementation details. The corresponding gait embeddings were structured into fixed-length sequences per penalty clip for each gait backbone under evaluation. To preserve class distribution, a consistent 80/20 train-test split was applied using stratified sampling based on the shot direction label. The same

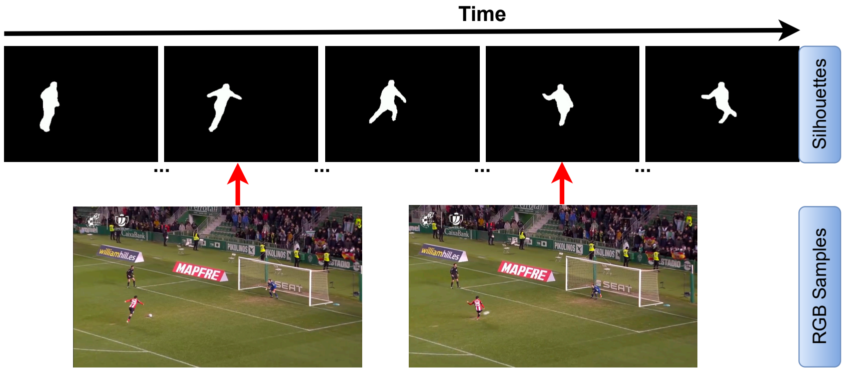

Figure 2: Silhouette sample frames from an RGB penalty sequence showing the kicker during the run-up and kick. The silhouette view is cropped to include only the kicker, excluding the goalkeeper, to focus on the kicker's motion patterns.

Table 1: Performance comparison of gait models and baselines for penalty kick direction classification. Accuracy and weighted recall are equal in this setting due to the use of single-label, multi-class classification with one prediction per sample. The goalkeeper baseline (human decision) is highlighted in gray, while the computational baseline (GaitBase) is highlighted in blue.

Model	**Pretrained Dataset**	**Accuracy**	**Precision**	**F1-Score**
Goalkeeper	N/A	46.0%	N/A	N/A
GaitBase (Fan et al., 2023)	OUMVLP	48.3%	45.2%	46.4%
GaitBase (Fan et al., 2023)	CASIA-B	51.7%	50.7%	51.1%
GaitPart (Fan et al., 2020)	OUMVLP	51.7%	54.3%	51.1%
GaitPart (Fan et al., 2020)	GREW	57.5%	**57.3%**	**57.3%**
GaitSet (Chao et al., 2018)	OUMVLP	50.6%	50.8%	50.6%
GaitSet (Chao et al., 2018)	CASIA-B	54.0%	52.4%	52.6%
GaitSet (Chao et al., 2018)	GREW	52.9%	54.8%	52.4%
GLN Phase 1 (Hou et al., 2020)	CASIA-B	56.3%	55.9%	55.5%
GLN Phase 2 (Hou et al., 2020)	CASIA-B	52.9%	52.1%	52.0%
GaitGL (Lin et al., 2021)	OUMVLP	54.0%	54.9%	53.9%
GaitGL (Lin et al., 2021)	CASIA-B	56.3%	52.9%	53.7%
GaitGL (Lin et al., 2021)	GREW	**58.6%**	49.9%	53.9%

split was reused across all models to ensure reproducibility, with index mappings stored and reloaded as needed. To account for variability due to model initialization and training dynamics, each experiment was repeated five times, and the resulting performance metrics were averaged to provide a robust estimate of model effectiveness.

The classification network was trained for up to 200 epochs with early stopping, using mini-batch size 32 and the Adam optimizer with default learning rate parameters. Class imbalance in the labels was addressed through weighted loss computation, using class weights derived from label frequency. As described in Section 3, the model architecture consisted of a single LSTM layer with 32 hidden units, followed by a dropout layer (rate 0.5), a ReLU-activated dense layer of size 16, and a softmax output layer with three units corresponding to the directional classes. Performance was measured using accuracy, precision, and F1-Score. Accuracy and weighted recall give the same result here because the model makes one prediction for each video, and each video has only one correct answer. Both metrics measure the same since we count how many predictions are correct overall.

Baselines. Two types of baselines were considered to contextualize the performance of gait-based models. The first is a human decision baseline derived from the goalkeeper's initial dive direction. Importantly, this does not indicate whether the goalkeeper successfully stops the shot. Instead, it simply reflects the goalkeeper's direction to dive, as recorded in the dataset. In some cases, goalkeepers may initiate their dive after the ball has already been struck. Therefore, this baseline can be seen as a best-case scenario for human anticipation, assuming access to all available cues before the shot is taken.

In addition to the human benchmark, a computational baseline was established using a lightweight gait recognition model commonly referred to as GaitBase. This architecture was selected due to its simplicity, reproducibility, and solid performance across controlled and unconstrained environments. Prior work has shown that even minimalistic gait encoders can yield competitive representations (Fan et al., 2023), making GaitBase a strong reference point for evaluating the added value of more sophisticated backbone designs in the context of penalty kick analysis.

5 EXPERIMENTAL EVALUATION

Evaluating various gait recognition models for penalty kick direction classification reveals interesting patterns across training datasets and model architectures (see Table 1). Notably, the highest-performing model overall was GaitGL trained on GREW, achieving an accuracy of 58.6% and an F1 score of 53.9%. This suggests that models trained on more diverse and in-the-wild datasets like GREW may better generalize to the natural variability present in broadcast soccer footage. GREW likely captures broader pose, scale, and environmental variation compared to more controlled datasets like OUMVLP or CASIA-B, which may contribute to its superior transfer performance.

Across the board, models trained on CASIA-B tended to perform more consistently than those trained on OUMVLP. CASIA-B's multi-view setup and moderate variability might offer a balance between structured feature learning and generalization to new domains. For example, GLN Phase 1 (trained on CASIA-B) achieved a strong F1 score of 55.5%, rivaling GaitGL's performance on GREW. GaitGL and GaitSet trained on CASIA-B also performed reliably, with F1 scores of 53.7% and 52.6% respectively. This trend highlights that while CASIA-B is a more controlled dataset, its design may still support domain transfer reasonably well for motion-based tasks.

In contrast, models trained on OUMVLP consistently yielded lower performance, despite the dataset's large scale. GaitPart, GaitSet, and GaitGL trained on OUMVLP all clustered around mid-range F1 scores, achieving 51.1%, 50.6%, and 53.9% respectively. The baseline model trained on OUMVLP lagged behind at 46.4% F1, suggesting that size alone does not guarantee effective transfer. OUMVLP's treadmill-based gait recordings may lack the naturalistic movement patterns found in soccer approach runs, reducing their representational relevance for this task.

In terms of design, GaitGL emerged as one of the most robust backbones across different training domains. It achieved top-tier results on GREW and CASIA-B and held up respectably on OUMVLP. Its performance consistency indicates a strong capacity for encoding discriminative motion dynamics in downstream classification tasks. GaitPart, although effective in surveillance-style gait recognition, performed moderately in this setup, likely due to its local part-based focus, which might miss out on full-body motion nuances relevant to shot prediction.

GLN also showed promise, especially in its Phase 1 variant trained on CASIA-B. It matched GaitGL in accuracy and demonstrated solid precision and recall values, supporting its suitability for fine-grained action prediction. Phase 1 corresponds to an intermediate checkpoint in the training of GLN, where the backbone is already capable of extracting meaningful gait representations but has not yet undergone full optimization for identity recognition. Interestingly, the Phase 2 variant, representing the final stage of training, underperformed slightly, with an F1 score of 52.0%. This performance drop may suggest that the additional training in Phase 2 biases the model more toward identity-specific features, potentially at the expense of general motion cues relevant to action understanding. Meanwhile, GaitSet exhibited a balanced yet unremarkable profile across all datasets, suggesting that while effective, its aggregation-based design may lack the temporal expressiveness needed for predicting dynamic actions like kicks.

Lastly, the baseline models served as important reference points. The CASIA-B-trained baseline outperformed the OUMVLP version by a notable margin (F1: 51.1% vs. 46.4%), reinforcing that training data characteristics critically shape downstream performance. While none of the baselines matched the top-performing models, their inclusion is crucial for interpreting the added value of more complex architectures. The results demonstrate that advanced gait models, especially GaitGL, GaitPart, and GLN, can extract semantically meaningful motion features for predictive tasks beyond identity recognition.

Error Analysis. Figure 3 presents the confusion matrices for GaitGL and GaitPart trained on the GREW dataset, revealing distinct misclassification patterns despite their similar overall accuracies (58.6% for GaitGL and 57.5% for GaitPart). GaitGL, while slightly more accurate, exhibits a pronounced bias toward predicting the *Right* class. Notably, it misclassifies 69.2% of actual *Center* kicks and 40.6% of *Right* kicks as *Right*, failing entirely to predict the *Center* class. This indicates potential overfitting to directional features dominant in right-sided kicks, possibly stemming from imbalances in pose or silhouette orientation during the run-up phase. Moreover, center kicks exhibit less exaggerated lateral body motion and

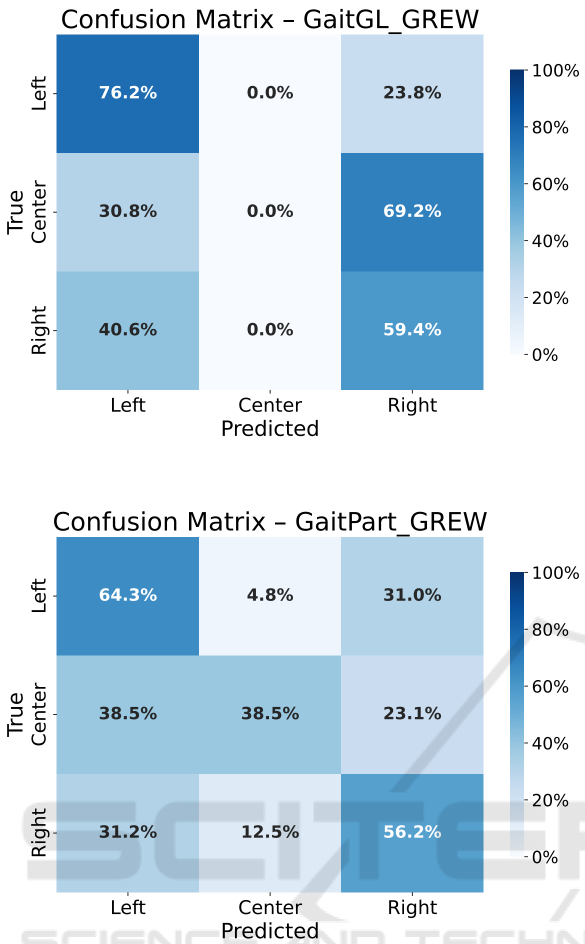

Figure 3: Normalized confusion matrices for models trained on GREW. Each matrix shows prediction performance across left, center, and right shot directions.

subtler preparatory cues than left or right shots. These traits may lead GaitGL to underrepresent or overlook the fine-grained temporal features critical for distinguishing center shots, especially in penalty scenarios where deceptive uniformity is typical.

In contrast, GaitPart demonstrates more balanced predictions across all three classes. While its confusion matrix still reflects a tendency to overpredict *Right* for ambiguous cases, it is the only model among the two that assigns non-zero predictions to the *Center* class (38.5% for true center kicks). Additionally, GaitPart maintains a better trade-off between precision and recall (57.3% and 57.5%), indicating that it is less prone to skewed predictions and captures a broader range of gait dynamics. This is particularly evident in its handling of true *Right* kicks, where it correctly classifies 56.2%, compared to 59.4% by GaitGL, but with fewer extreme misclassifications.

These observations suggest that while GaitGL may achieve marginally higher accuracy by confidently predicting dominant classes, it lacks nuance in handling more ambiguous gait patterns, especially those associated with central shot directions. GaitPart, despite its slightly lower accuracy, appears to better generalize across class boundaries. This trade-off highlights an important consideration: models with higher overall accuracy may still perform poorly on minority or difficult-to-classify cases, and confusion matrix analysis is essential for understanding these hidden weaknesses.

6 CONCLUSIONS

This study explored gait-based representations for predicting penalty kick direction in soccer, extending gait recognition beyond its usual role in identity verification. We used pre-trained gait backbones and a lightweight temporal classifier to assess whether motion patterns during the run-up could reliably indicate shot direction. To support this, we curated a new dataset of broadcast penalty sequences and evaluated the transferability of gait embeddings across various architectures and training domains.

Results showed that both the choice of backbone and training data significantly impact performance. Models trained on GREW, an in-the-wild dataset, performed best overall, with GaitGL achieving the highest accuracy and F1 score. However, this came with strong class bias, particularly toward right-side predictions. GaitPart, though slightly less accurate, yielded more balanced results across all directions, including the underrepresented center. This highlights the need to go beyond global metrics and consider per-class behavior in imbalanced tasks.

These findings suggest that gait encodes meaningful information about players' motor intentions, even in dynamic, high-pressure contexts like penalty kicks. Our approach builds on recent work using Human Action Recognition (HAR) models for penalty analysis (Freire-Obregón et al., 2025), which focus on cropped segments like the run-up or kick. In contrast, by preserving the entire sequence, our model captures a broader range of kinematic and contextual cues. While this may limit some traditional gait applications, such as short-cycle identification, it enables exploration of richer temporal patterns. Notably, despite departing from standard gait usage, the results show that extended motion cues carry predictive value. This opens the door to personalized modeling of penalty direction, potentially tailored to each player's unique movement signature.

This research paves the way for practical sports applications, such as decision-support tools, player diagnostics, and automated video analysis. Incorporating features like foot orientation, ball trajectory, or goalkeeper behavior could boost accuracy and contextual understanding. Future work may explore multi-modal fusion or adapt gait models to better capture sport-specific movement patterns.

ACKNOWLEDGEMENTS

This work is partially funded funded by project PID2021-122402OB-C22/MICIU/AEI /10.13039/501100011033 FEDER, UE and by the ACIISI-Gobierno de Canarias and European FEDER funds under project ULPGC Facilities Net and Grant EIS 2021 04.

REFERENCES

Aharon, N., Orfaig, R., and Bobrovsky, B.-Z. (2022). BoT-SORT: Robust associations multi-pedestrian tracking. *arXiv preprint arXiv:2206.14651*.

Artiles, J., Hernández-Sosa, D., Santana, O., Lorenzo-Navarro, J., and Freire-Obregón, D. (2024). Classifying soccer ball-on-goal position through kicker shooting action. In *Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods - ICPRAM*, pages 79–90. INSTICC, SciTePress.

Chao, H., He, Y., Zhang, J., and Feng, J. (2018). Gaitset: Regarding gait as a set for cross-view gait recognition. In *AAAI Conference on Artificial Intelligence*.

ChyronHego (2017). 5 reasons you need tracab player tracking. Accessed 12 Feb 2017.

Fan, C., Liang, J., Shen, C., Hou, S., Huang, Y., and Yu, S. (2023). Opengait: Revisiting gait recognition towards better practicality. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Fan, C., Peng, Y., Cao, C., Liu, X., Hou, S., Chi, J., Huang, Y., Li, Q., and He, Z. (2020). Gaitpart: Temporal part-based model for gait recognition. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14213–14221.

Figueroa, P., Leite, N., and Barros, R. (2006). Tracking soccer players aiming their kinematical motion analysis. *Computer Vision and Image Understanding*, 101:122–135.

Freire-Obregón, D., Lorenzo-Navarro, J., Santana, O. J., Hernández-Sosa, D., and Castrillón-Santana, M. (2022). Towards cumulative race time regression in sports: I3d convnet transfer learning in ultra-distance running events. In *2022 26th International Conference on Pattern Recognition (ICPR)*, pages 805–811.

Freire-Obregón, D., Santana, O. J., Lorenzo-Navarro, J., Hernández-Sosa, D., and Castrillón-Santana, M. (2025). Predicting Soccer Penalty Kick Direction Using Human Action Recognition. In *2025 23rd International Conference on Image Analysis and Processing (ICIAP)*.

Hou, S., Cao, C., Liu, X., and Huang, Y. (2020). Gait Lateral Network: Learning Discriminative and Compact Representations for Gait Recognition. In *European Conference on Computer Vision*, page 382–398.

Innovations, H.-E. (2017). Hawk-eye goal line technology. Accessed 12 Feb 2017.

Jocher, G., Chaurasia, A., and Qiu, J. (2023). Ultralytics yolov8.

Kapela, R., McGuinness, K., Swietlicka, A., and O'Connor, N. (2014). Real-time event detection in field sport videos. In Moeslund, T., Thomas, G., and Hilton, A., editors, *Computer Vision in Sports*, chapter 14. Springer.

Lin, B., Zhang, S., and Yu, X. (2021). Gait recognition via effective global-local feature representation and local temporal aggregation. *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14628–14636.

Liu, J. and Carr, P. (2014). Detecting and tracking sports players with random forests and context-conditioned motion models. In Moeslund, T., Thomas, G., and Hilton, A., editors, *Computer Vision in Sports*, chapter 6. Springer.

Maksai, A., Wang, X., and Fua, P. (2016). What players do with the ball: A physically constrained interaction modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Manafifard, M., Ebadi, H., and Moghaddam, H. (2016). A survey on player tracking in soccer videos. *Computer Vision and Image Understanding*. Submitted for Special Issue on Computer Vision in Sports.

Takemura, N., Makihara, Y., Muramatsu, D., Echigo, T., and Yagi, Y. (2018). Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSJ Transactions on Computer Vision and Applications*, 10:1–14.

Thomas, G., Gade, R., Moeslund, T. B., Carr, P., and Hilton, A. (2017). Computer vision for sports: Current applications and research topics. *Computer Vision and Image Understanding*, 159:3–18.

Yang, C.-Y., Huang, H.-W., Chai, W., Jiang, Z., and Hwang, J.-N. (2024). SAMURAI: Adapting segment anything model for zero-shot visual tracking with motion-aware memory.

Yu, S., Tan, D., and Tan, T. (2006). A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. *18th International Conference on Pattern Recognition (ICPR'06)*, 4:441–444.

Zhu, Z., Guo, X., Yang, T., Huang, J., Deng, J., Huang, G., Du, D., Lu, J., and Zhou, J. (2021). Gait recognition in the wild: A benchmark. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14769–14779.