

Less is More: End-to-end Learning of Insights from a Single Motion Sensor for Accurate and Explainable Soccer Goalkeeper Kinematics

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

20-04-2021 / 22-04-2021

CITATION

Lisca, Gheorghe; Prodaniuc, Cristian; Grauschoff, Thoms; Axenie, Cristian (2021): Less is More: End-to-end Learning of Insights from a Single Motion Sensor for Accurate and Explainable Soccer Goalkeeper Kinematics. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.14452707.v1>

DOI

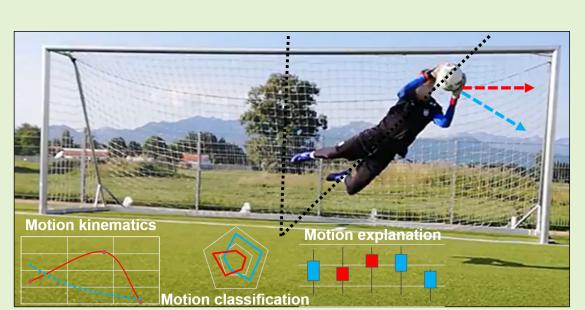
[10.36227/techrxiv.14452707.v1](https://doi.org/10.36227/techrxiv.14452707.v1)

Less is More: End-to-end Learning of Insights from a Single Motion Sensor for Accurate and Explainable Soccer Goalkeeper Kinematics

Gheorghe Lisca, Cristian Prodaniuc, Thomas Grauschoff, and Cristian Axenie, *Member, IEEE*

Abstract—On-field sensor-based soccer player tracking solutions are emerging and provide new insights into the dynamics of the player during training or a match. Yet, not all player positions are equally privileged. Goalkeepers' training and performance assessment were for a long time ignored. Understanding what is "the side of the post where most high dives were performed" provides valuable input for both the trainer and the athlete to improve performance or avoid injuries. In the current study, we focus on a practical methodology to extract insights from goalkeeper kinematics to inform such analytics. We demonstrate that information from a single motion sensor can be successfully used for learning patterns in goalkeeper's motion and provide an explainable goalkeeper kinematics assessment. We employed raw and quaternions data and we evaluated a series of machine learning algorithms that discriminate dive types (i.e. binary classification) and dives from other types of specific motions (i.e. multi-class classification) directly from the data. Our results demonstrate that XGBoost outperforms other approaches in terms of performance when considering both raw and quaternions, essentially benefiting from both types of data. Additionally, each prediction of the model is accompanied by an explanation of how each sensed motion component contributes to describing a specific goalkeeper's action captured by the model. The explainable predictions along with the efficient deployment of XGBoost were decisive in our applied study. We evaluated our methodology on a first batch of experiments using online available data from 7 goalkeepers during 30 minutes-long training sessions.

Index Terms—Soccer Goalkeeper, Embedded Sensors, Machine Learning, Motion Classification.



I. INTRODUCTION

Arguably, the goalkeeper is the most important position on the soccer field. With one brilliant save a goalkeeper can win a game or become the "persona non grata" with an ill-timed dive. Agility, speed, and fitness are required for the nature of the position [1]. For decades, goalkeepers were expected to train alone or were just simply incorporated into routine practice sessions, with very little consideration being paid to the unique needs of the goalkeepers.

Goalkeepers are clearly different from players on the field and must be trained accordingly [2]–[4]. The first step towards effective training is accurate motion assessment. Such an assessment has value when it provides a quantifiable measure

Gheorghe Lisca is with the Technische Hochschule Ingolstadt, Esplanade 10, 85049, Ingolstadt, Germany (e-mail: gheorghe.lisca@thi.de).

Cristian Prodaniuc is with Goalplay GmbH - Co. KG, Südliche Münchner Straße 30, 82031 Grünwald, Germany (e-mail: cristian.prodaniuc@googlemail.com).

Thomas Grauschoff is with the Technische Hochschule Ingolstadt, Esplanade 10, 85049, Ingolstadt, Germany (e-mail: thomas.grauschoff@thi.de).

Cristian Axenie is with the Audi Konfuzius-Institut Ingolstadt Lab at the Technische Hochschule Ingolstadt, Esplanade 10, 85049, Ingolstadt, Germany. (e-mail: cristian.axenie@audi-konfuzius-institut-ingolstadt.de).

of how each motion component contributes to explaining a goalkeeper's kinematics during his motion [5], [6], for instance anticipate dive direction from the kinematics of a particular body segments [7]. The extracted insights can be used in designing injury avoiding strategies [6] or improve saving performance [8]–[10] by building a goalkeeper's biomechanical profile [11].

A. Sensor-based kinematics assessment

Typically, a kinematic assessment is performed in a controlled lab environment [6], [11]–[13], but there are also initial deployments on the field [8], [14], [15]. Opposite to the lab assessment, which involves a large amount of sensors (e.g. cameras with body-worn tracking markers [16], large number of body-worn inertial sensors [17]–[19]), controlled experiments, and precise calibration, on-field deployments use only a limited number of inertial sensors [18], [20]. Such sensors capture the fast changes in the motion of the goalkeeper in a realistic scenario, but are poised by inherent noise [4], [21] and inter-player diving variability [22], [23]. The sensor-based motion assessment captures information about the athlete and his context using various motion-generated signals [24], such as, muscles properties [25] and center of

gravity [10] to describe performance [1], [23]. Such sensory data is typically composed to evaluate the diving ability, as derived from ground reaction force data, and, for instance, anthropometry [13]. Finally, in a study comparatively examining fast hand movements and orientations, [26] introduced an informative evaluation of various machine learning (i.e RBF SVM, 2D - CNN, and LSTM) when using raw data and Euler angles. Despite the good evaluation, the study was performed in constrained conditions using self-collected sensory data samples, highly biased and incapable to capture intra-player variability.

B. Data processing

Most sensor-based soccer motion recognition applications developed so far are described as series of sequential computations, known as activity recognition chains. These chains are responsible for extracting information from sensor signals, such as peak detection, segmentation, and use it for classification [27]–[29]. Multiple such methods to segment sensor signals, extract features and to select, rank, and classify features have been studied [28], [30]–[32], but each one was limited to a certain aspect of motion.

Despite the broad range of sensing equipment and data processing techniques [24] little focus has been given to goalkeepers' kinematics extraction [33]. Most of the existing solutions target tracking and on-field stats, including kicks, sprints, travelled distance, and max speed [20], [34]. Such on-field approaches use a limited number of sensors and either focus on modeling the impact and kick detection [35] or advanced algorithms for classification of sequential movements in training [36]. Recently, the community has turned its attention to the so-called end-to-end pipelines for motion recognition [37]–[40]. Such methods, use advanced machine learning algorithms to extract insights from wearable sensors data in order to learn body kinematics [38], [39] or kinematics correlation to other bio-signals [41].

C. Motivation of the study

As the state-of-the-art shows, soccer goalkeepers are a category typically neglected in kinematics assessment studies. This is due to the fact that there is always a trade-off to be made between: 1) the accuracy of lab assessment [6], [12] and the unreliability and variability of the on-field assessment [14], [15]; 2) the number and type of sensors used for the assessment [16], [18]; 3) the measured kinematic quantities [1], [10], [25], [42]; 4) the type of data features used in the analysis [28], [29] and their capability to capture a complete motion profile [11], [36]; 5) the complexity and parametrization of the data processing algorithms [39], [40] versus their explainability; and, finally, 6) the amount of resources invested in goalkeeper's training.

At the top of the solution hierarchy, machine learning methods exploit directly the probabilistic structure and peculiarities of raw data [14], [38], [40], [42], thus, alleviating the need to model data and, eventually, hand-craft heuristics which will only be limited in applicability [13], [25]. Our motivation is

to use machine learning to understand goalkeeper's kinematics in order to materialize complex analytics. For instance, understanding what is "the side of the post where most dives were performed" provide valuable input both the trainer and the athlete to improve performance or avoid injuries. Such a dive type histogram analysis describes the "semantics" of the physical burden of the joints and muscles of the diving sides of the goalkeeper [6]. For instance, goalkeepers performing a rolling motion at dive initiation reduce their hip loading. Such biomechanics insights can contribute to the establishment of injury criteria and avoidance measures [43].

D. Objectives of the study

Our study addresses the trade-offs and constraints that the problem of extracting accurate goalkeeper kinematics poses. It demonstrates the advantages that data-driven machine learning approaches bring, even from a small number of sensors. The experimental data was collected from sensory modules installed in the goalkeeper gloves under two forms: raw data (i.e. 3D acceleration, angular velocity, magnetic field intensity) and quaternions. The objective of our study is threefold:

- to demonstrate that a single motion sensor data is sufficient to accurately classify goalkeeper motions;
- to demonstrate that machine learning algorithms have the potential to explain how each sensed motion component contributes to describing a specific goalkeeper action;
- to propose an end-to-end approach for kinematics assessment with high potential for real-world deployment, offering:
 - cheap and unobtrusive, limited sensing equipment,
 - accurate and explainable decision making based solely on limited sensory data,
 - the potential for providing personalized feedback

This system is meant and practically deployed to perform an assessment of the training session which is typically designed by a trainer. Note that, the trainer prepares the exercise plan and the system provides a quantitative and qualitative assessment of the training outcome.

II. MATERIALS AND METHODS

In the current section, we introduce the experimental setup, data acquisition, and the data processing methods we employed in our study.

A. Hardware

In all our experiments, we used data from 7 goalkeepers using a sensing device embedded in a pair of Goalplay¹ goalkeeper gloves (Figure 1). The data acquisition system was a MicroHub development kit². The modular platform for wearable applications included a 9-axis Inertial Measurement Unit (IMU) (3D accelerometer, 3D gyroscope, and a 3D magnetometer), a communication module (i.e. Bluetooth Low Energy (BLE)), and a memory storage (i.e. a 2GB SD card).

¹<https://www.goalplay.com/en>

²<http://www.interactive-wear.com/r-and-d>

The embedded accelerometer outputs acceleration in units of gravity g within ± 2 g and ± 16 g. The gyroscope outputs angular velocity in degrees per second (dps) within ± 250 dps and ± 2000 dps. The magnetometer outputs magnetic field intensity values (uH) proportional with the absolute angle values (deg) - offset to magnetic north. All the sensors in the glove module were parametrized for appropriate sensitivity for the task, i.e. ± 2 g acceleration range and ± 250 dps gyroscope range. To avoid communication problems during



Fig. 1. Sensor equipped goalkeeper gloves.

the lengthy exercise sessions, data from the embedded sensors was acquired with a 16-bit resolution at a frequency of 100 Hz and stored on the embedded SD card with a capacity of 2 GB. In order to estimate angles from raw accelerometer we elaborated the mathematics of the reference frame rotations and used geometrical reflections to extract the nonlinear relations depending on the rotated gravitational acceleration (i.e. \arctan of ratio between X, Y and Z-axis components of acceleration). The gyroscope on-board the gloves module measured angular velocities, so we numerically integrated the raw data using a simple sum approximation (explicit Euler method with step width dt). To obtain the magnetometer readings mapping from raw magnetic field intensity to angles, we described the rotations between the object frame (the gloves) and the reference frame to estimate the rotated earth magnetic field with respect to the gloves reference frame and geometrical reflections (i.e. \arctan of ratio between X and Y-axis components of magnetic field).

B. Data collection

Aside the raw sensory data from the accelerometer (i.e. acceleration on x, y, z axes and longitudinal acceleration), gyroscope (i.e. angular velocity on x, y, z axes), and magnetometer (i.e. magnetic field intensity on x, y, z axes), the embedded sensor provides also quaternions. The choice to represent such a compound motion through raw and quaternions data is meant capture the typical "signatures" of the broad range of motions and explosive changes in goalkeeper's actions, typically visible in raw data patterns. Quaternions, compared to Euler angles (i.e. roll, pitch, and yaw) are simpler to compose and avoid the gimbal lock (i.e. problems with measuring orientation when the pitch angle approaches +/- 90 degrees). Moreover,

compared to rotation matrices, quaternions offer a more compact and numerically stable representation. But, as we know quaternions have limitations. Compared to affine transformation matrices, quaternions only contain a rotation, and no translation and scaling. This makes it more difficult to check whether two quaternions represent the same rotation by simply checking their components for equality. This is where the raw data can be used to disambiguate the motion components involved in a certain dive. As one of the goals of the study is to provide an end-to-end approach for kinematics assessment, we evaluated the use of raw data only, quaternions only, and the combination of raw and quaternions data, respectively. The reason is to understand what are the most informative representation cues for precise and explainable goalkeeper's motion profile (see Results section).

The goalkeepers data was temporally segmented by experts in order to identify (over pre-defined time windows) the possible exercises in the stream of motion data from the sensory module. In order to perform the segmentation the video recording and glove sensory data was aligned. Subsequently, temporal segments (i.e. templates) were extracted offline describing a certain exercise. This template was then compared (with a parametrized variance) with the incoming stream.

System calibration was performed against the ground truth camera data. Unit calibration of the sensors inside the glove module was performed by comparing the data against the recorded videos for each of the goalkeepers for predefined initial positions or events (i.e. hand clapping, ball bouncing - given the data segmentation we performed). Value calibration was performed by comparing the raw output of the sensors with the analytical formulation of derived quantities (see Hardware). We additionally performed a leave-one-out cross-validation and demonstrated that the system doesn't require to be calibrated to each goalkeeper separately (see Table III and Figure 5).

The datasets contain recorded data from 7 male, right-handed, 10 to 17 years old, goalkeeper trainees with 1 to 7 years of experience, and were 1.5 to 2.01 meters tall. Before each 30 minute-long session, each goalkeeper warmed up, stretched, and then started the exercise execution. The datasets are publicly available [44] on the Zenodo platform³.

C. Data annotation

As ground truth for the experiments, synchronized video recordings of the training sessions were captured. The Goalplay team annotated every repetition of the training exercises (i.e. dives, catches, throws) in both video and matching embedded glove sensor data, as shown Figure 2. The segmentation process was applied on this labelled dataset. Additionally, besides the exercises, other motions of the goalkeepers performed during the training, such as back passes, sprints, hand claps, and ball bouncing, were annotated. This was done to ensure a proper (i.e. expert) segmentation of the motion data. In order to synchronize the data from the glove sensor and the video capture, the goalkeepers clapped three times in

³<https://zenodo.org/record/4629325>

the beginning, after each set of exercises, and at the end of the training session. The annotations, shown in Table I, were done by multiple Goalplay experts and were reviewed by one of the authors to ensure consistency. In total, the datasets contained, for each of the 7 goalkeepers, around 2100 labeled motion instances, out of which 1100 were actual training exercises and 1000 instances of other motion, such as passes and hand claps, respectively. The problem we consider in this study faces a clear data imbalance. The valid exercises, extracted using the segmentation, account for 35% of the overall data. We tackled this problem in our experiments, as described in the Experiments section.

D. End-to-end Machine Learning

Machine learning algorithms are useful tools for tackling the automatic extraction of features from large sets of data and, more precisely, the extraction of unique features characterising different classes of similar features, typically found in motion kinematics [42], [45], [46] and motion classification [14], [47]. When considering motion understanding and kinematics extraction from limited (inertial) sensing, only a few approaches stand out [37]–[39], and only a very small number apply to soccer goalkeeping [14], [42].

Being guided by practical reasons and given the nature of the problem and the data, the approach we proposed doesn't promote the use of Deep Learning. Such approaches are based on affine projections of input signals with outlier removals in their activation layers (i.e. basically behaving as regions clippers). In principle, the separation function gets adapted to a larger region in input space. Based on this observation, such approaches do not really find a solution but rather compare the current input with previous observations, so that the actual recognition job was done by the GoalPlay team segmenting the data. Our contribution revolves around the idea of combining raw and quaternions data to augment the motion components describing the multiple "class templates". Given the supervised nature of the problem, traditional Machine Learning models were chosen to Deep Learning. In particular, considering a Long-Short-Term-Memory (LSTM) would have been advantageous when the dataset would not have the expert labelling and temporal segmentation available. Basically, by learning sequences in the raw glove sensors timeseries, the LSTM could capture and better preserve the "long-range dependencies" in the multiple raw sensory covariates [48]. Another possible approach, the 1D Convolutional Neural Networks (1D CNN) learns to extract features from the sequences of sensory observations and maps these features to different goalkeeper types [49]. The benefit of using CNNs for motion classification from raw data would have been prominent when no domain expertise to manually engineer input features was available.

In our study, we developed an end-to-end approach to extract goalkeeper kinematics from a single embedded sensors module (i.e MicroHub, see Hardware section) based on a Tree Ensemble Method, namely the Extreme Gradient Boosting (XGBoost) [50]. XGBoost is a highly effective and widely used tree boosting method. In XGBoost the multiple learners, the trees, are built sequentially such that each subsequent tree

aims to reduce the errors of the previous tree during training. Each tree in the ensemble learns from its predecessors and updates the prediction mismatch. Hence, the tree that grows next in the sequence will learn from an updated version of the errors. Each of these weak learners (i.e. high bias and variance) contributes to the prediction, enabling XGBoost to produce a strong learner (i.e. low bias and variance) by effectively combining these weak learners.

Our choice of the XGBoost ensemble model is motivated by the following aspects:

- invariance to scaling of inputs, so features normalization is secondary, allowing for native use of raw (inertial) data;
- learning higher-order interactions between raw data (or features), without expensive and bias-prone feature engineering;
- employing regression trees, which perform well for both classification and regression at scale;
- using an explainable learning structure due to its tree structure;
- deployable for inference in the embedded glove system [51]

In our implementation, in order to parametrize the XGBoost ensemble, we ran a grid search for minimizing the Receiver Operating Characteristic / Area Under Curve (ROC / AUC) together with a 10-fold stratified cross-validation on 25% test and 75% training dataset split. In our validation approach, we used stratified k-fold cross-validation to balance the dataset by ensuring that each class represents approximately in equal proportion in all the sample and, hence, reduce the variance of the estimates. Additionally, we use Leave-One-Out Cross-Validation (LOOCV) to demonstrate that the system doesn't require to be calibrated to each goalkeeper separately (see Results). The best performing XGBoost ensemble configuration, used in all the experiments, employed: 50 trees, 0.5 decreasing learning rate to avoid overfitting, 3 as maximum depth of a tree, a logistic function optimization objective, and L2 regularization term on weights to make the model more conservative. For the performance evaluation, we considered the traditional metrics: Accuracy, Precision, Recall, and F1 score, respectively. Additionally, in order to interpret the impact of having a certain value for a given raw data feature in comparison to the prediction the model would make if that feature took some baseline value, we used SHAP (SHapley Additive exPlanations) [52]. More precisely, we employed a variant of SHAP, TreeSHAP, for tree-based machine learning models [53]. Basically, the TreeSHAP explainer uses the conditional expectation to capture the influence on the prediction function of a feature also when the feature is correlated with another feature that has an influence on the prediction.

E. Experiments

In order to explore the capabilities of our end-to-end learning system, we employed an online dataset containing a series of experiments that exploit the limited sensing capabilities of a single motion sensor to classify and explain a goalkeeper's motion. In our experiments, we collected data from 7 young goalkeepers during their training sessions with a professional

| Exercise | Labels | Repetitions |
|---------------|---|----------------|
| Dive | diveLowRight, diveHighRight, shortDiveRight, diveLowLeft, diveHighLeft, shortDiveLeft | 5 |
| Catch | catchHand, catchBody, catchGround | 10 |
| Dive Stand | diveStandLowRight, diveStandHighRight, diveStandLowLeft, diveStandHighLeft | 5 |
| Throw | throwHigh, throwLow | 10 |
| Jump Catch | jumpCatchStand, jumpCatchRun | 8 |
| Other motions | pass, shortSprint, longSprint, jogging, joggingJumping, kickBody, kickGround, bounce, clap, ballPickUp, armRotation | 65% of session |

TABLE I

GOALKEEPER EXERCISES AND LABELS SUMMARY. EACH LABEL DESCRIBES A CERTAIN TIME WINDOW OF INERTIAL DATA AND ITS MATCHING NUMBER OF FRAMES ACQUIRED WITH THE CAMERA THAT DESCRIBE A CERTAIN MOTION. A SAMPLE OF EXERCISES IS DEPICTED IN FIGURE 3.

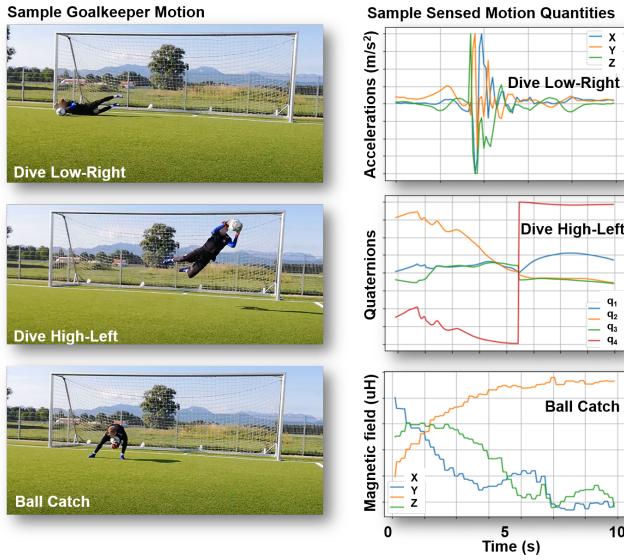
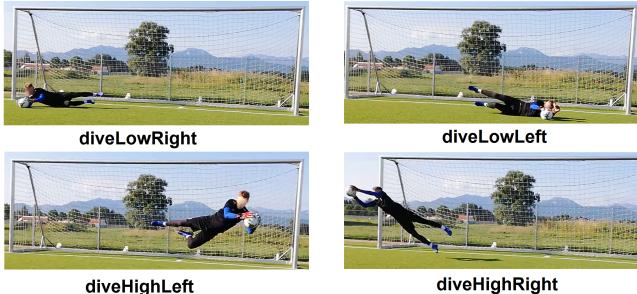


Fig. 2. Sample goalkeeper motion and available sensory data.

Sample dives



Sample catches

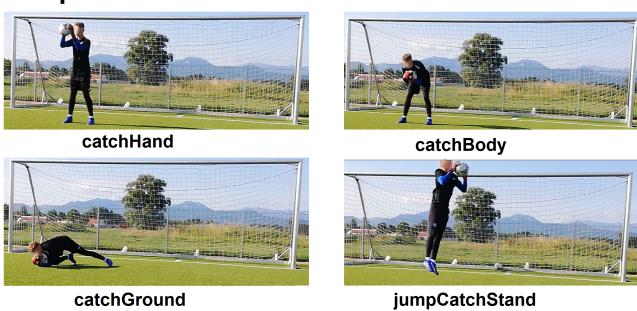


Fig. 3. Sample exercises: dives and catches.

goalkeeper trainer from Goalplay. During the average 33 minutes session, the 7 goalkeepers performed different exercises including variations of dives, throws, and catches, as shown in Table I. Each type of exercise assessment was relative to the post. For instance, there were 4 types of dives, 2 heights (i.e. low and high) for each side of the post. The trainer adjusted the angle and intensity of the balls sent to the post. After each shoot, goalkeepers passed the ball back to the trainer and signaled ready for the next repetition (i.e. for data synchronization purposes). The data acquired during the series of exercises described in Table I were batched in a per session, per goalkeeper dataset and was subsequently fed to the end-to-end learning system. The analysis code to reproduce the figures in the paper is available on GITLAB⁴. We started our analysis by exploring a simple task, discriminating goalkeeper motion using a binary classification model (i.e. dive or no dive), independent of the side or the level (i.e low left, high right). In the second part, we evaluated the model on a multi-class setup, where the system received data with labels for all types of motion the goalkeeper performed, as shown in Table I. Important to mention is the fact that in such a scenario the distribution of data examples across the classes is skewed. Basically, a typical exercise session (see Table I), contains predominantly less dives as other exercises or motion types. This describes a typical class imbalance problem which poses a challenge for predictive modelling. In our approach, we designed the system without the assumption of an equal number of examples for each class (i.e. we have 26% dives in an exercise session). This results in the model having a better predictive performance, specifically for the minority class, without being more sensitive to classification errors for the minority class (i.e. dives) than the majority class (i.e. other motions). To cope with the imbalance, we used Adaptive synthetic sampling approach for imbalanced learning (ADASYN) [54], an extension of SMOTE, for creating more examples in the vicinity of the boundary between the two classes than in the interior of the minority class, for a 35% ratio minority-majority after resampling. In both binary and multi-class setups, we fed the system subsequently with: raw data only (R), quaternions data only (Q), and both raw and quaternions data (R&Q), respectively, as shown in Table II and Table III. In order to evaluate the performance

⁴<https://gitlab.com/akii-microlab/goalkeeper-motion-understanding-framework/>

of the system, we used typical classification metrics, namely Accuracy, Precision, Recall, and F1-Score, respectively. The experiments and the performance evaluation had the purpose to demonstrate that a single motion sensor data is sufficient to accurately classify goalkeeper motions. Additionally, in order to demonstrate that machine learning algorithms have the potential to explain how each sensed motion component contributes to describing a specific goalkeeper action, we evaluated the feature importance in each data and classification configuration. Finally, all experiments were complemented by a Shapley additive explanations analysis. This analysis provided an overview on how each feature contributes to the final prediction, along with samples of individual predictions in order to emphasize the explainable decision making of our system based solely on raw data.

III. RESULTS

In the current section, we introduce our results and touch several relevant points for the assessment of the goalkeeper kinematics of our end-to-end learning system.

Feature importance

In order to tune our model, we looked at the top 3 most important features for each of the experimental scenarios. The results of our analysis are given in Table 3. The F score in the feature importance computation of our system represents the fractional contribution of each feature to the model based on the total gain of this feature's splits in the model's regression trees (i.e. in our case we have 50 tree estimators with up to a depth of 3 and around 720 splits per level). Basically the role of the splits is to minimize the squared error greedily in order to fit to the gradient of the loss function. Finally, a higher score depicts a more important predictive feature, obtained by calculating the average gain across all splits where a feature was used.

Performance evaluation

The purpose of the performance evaluation was to identify the best training data configuration for which the end-to-end learning platform yields best performance in the two tasks (i.e. binary (dive) discrimination and multi-class classification).

In order to better motivate the choice of the ensemble model, we compared the performance of our system against four typically used machine learning models, namely: Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Quadratic Discriminant Analysis (QDA), respectively. All the models were trained for both binary and multi-class setups on all the possible data types (i.e. raw data only (R), quaternions data only (Q), and the combination of raw and quaternions data (R&Q)) and evaluated using k-Fold ($k=10$) stratified cross validation and LOOCV. The basic parametrization of the comparative models is as following:

- ANN: 2 layers, 100 neurons each, optimizing the log-loss function using stochastic gradient descent (ADAM), batch size of 200, 0.0001 L2 penalty regularization;
- SVM: 1.0 squared L2 penalty regularization, Radial Basis Function (RBF) kernel with sub-unit features coefficient

γ , returns the one-vs-rest decision function;⁵

- QDA: fitting class conditional densities (i.e. Gaussian) to the data and using Bayes' rule with per-class covariance estimates regularization.

Our approach detected over 93% of most exercise types and filtered out most of the irrelevant motions. Only 140 instances of irrelevant motion were wrongly detected as relevant exercises, which is a small fraction (14%) of the total amount of instances of irrelevant motion we annotated (1000) and an even smaller fraction of every irrelevant motion goalkeepers performed during the training session, which we did not annotate (e.g., walking, running, picking up balls from the ground, receiving balls passed at them without any specific technique).

As one can see in Table III, the system yields best performance when using the combination of raw and quaternions data. The second best performance is obtained when the system uses the raw data alone, whereas the scenario in which the system uses only the quaternions data yields an average performance. On one side, this happens due to the fact that the raw data captures the low-level kinematic parameters of goalkeeper's motion (i.e. accelerations, lateral accelerations, angular velocities). On the other side, the quaternions data is the transformation result of the low-level data into 3D motion bearing higher level semantics. This provides a gain in the global understanding of the motion, but doesn't capture the fast changes in kinematic parameters. Yet, the semantics of the quaternions data in combination with the raw data enable the system to accurately discriminate the dives and to correctly segment the other types of motion (i.e. multi-class classification task).

Important to note that in our study, accuracy is considered in the LOOCV as a performance metric facing imbalanced classification. In our performance evaluation we used updated metrics for imbalanced classification. Here, precision quantifies the number of positive class predictions that actually belong to the positive class. Recall quantifies the number of positive class predictions made out of all positive examples in the dataset. The Macro F1-Score provides a single score that balances both precision and recall in one number. In order to support the evaluation of the system for the two tasks, we present the confusion matrices for all the data configurations (i.e. raw data only (R), quaternions data only (Q), and the combination of raw and quaternions (R&Q)) for dive discrimination in Table III. This provides more insight in the performance of our system on the set of test data for which the true values are known (see Figure 4).

We complement the evaluation of the different approaches (i.e. XGBoost, MILP, SVM, and QDA) with a significance testing following the next procedure:

- Read relevant data from experiment results for each of the four approaches.
- Compute relevant traffic aggregation metrics (i.e. K-fold Cross-Validation: Precision, Recall, Macro F1-Score; and

⁵The key practical reason to chose this kernel type is that the penalty for the RBF network depends on the centers of the RBF network (and hence on the sample of data used) whereas for the RBF kernel, the induced feature space is the same regardless of the sample of data, so the penalty is a penalty on the function of the model, rather than on its parameterisation.

| | Binary discrimination | | | Multi-class classification | | |
|-------------|-----------------------|--------|--------|----------------------------|--------|--------|
| | Rank 1 | Rank 2 | Rank 3 | Rank 1 | Rank 2 | Rank 3 |
| Data | | | | | | |
| R | ax | my | lay | lay | laz | my |
| Q | q2 | q4 | q3 | q2 | q4 | q3 |
| R&Q | q4 | q2 | mz | laz | ay | lax |

TABLE II

FEATURE IMPORTANCE RANKING IN THE SYSTEM FOR EACH OF THE DATA TYPES FED TO THE SYSTEM: RAW DATA ONLY (R), QUATERNIONS DATA ONLY (Q), AND THE COMBINATION OF RAW AND QUATERNIONS DATA COMBINED (R&Q).

| | Binary discrimination | | | | Multi-class classification | | | |
|-------------------|-----------------------|-------------|----------------|-------------|----------------------------|-------------|----------------|-------------|
| | Precision | Recall | Macro F1-Score | Accuracy | Precision | Recall | Macro F1-Score | Accuracy |
| Model/Data | | | | | | | | |
| XGboost | | | | | | | | |
| R | 0.87 | 0.85 | 0.88 | 0.91 | 0.90 | 0.91 | 0.90 | 0.94 |
| Q | 0.74 | 0.70 | 0.77 | 0.93 | 0.83 | 0.86 | 0.84 | 0.90 |
| R&Q | 0.97 | 0.94 | 0.94 | 0.97 | 0.93 | 0.94 | 0.95 | 0.97 |
| ANN | | | | | | | | |
| R | 0.85 | 0.84 | 0.86 | 0.92 | 0.79 | 0.85 | 0.85 | 0.90 |
| Q | 0.69 | 0.54 | 0.67 | 0.94 | 0.74 | 0.83 | 0.78 | 0.87 |
| R&Q | 0.88 | 0.76 | 0.87 | 0.95 | 0.84 | 0.87 | 0.87 | 0.92 |
| SVM | | | | | | | | |
| R | 0.81 | 0.67 | 0.78 | 0.91 | 0.74 | 0.84 | 0.79 | 0.85 |
| Q | 0.69 | 0.43 | 0.59 | 0.93 | 0.69 | 0.83 | 0.78 | 0.86 |
| R&Q | 0.86 | 0.84 | 0.89 | 0.92 | 0.78 | 0.84 | 0.80 | 0.94 |
| QDA | | | | | | | | |
| R | 0.75 | 0.56 | 0.69 | 0.89 | 0.83 | 0.80 | 0.87 | 0.88 |
| Q | 0.53 | 0.27 | 0.39 | 0.78 | 0.71 | 0.82 | 0.79 | 0.85 |
| R&Q | 0.80 | 0.70 | 0.80 | 0.88 | 0.86 | 0.79 | 0.88 | 0.90 |

TABLE III

PERFORMANCE EVALUATION OF THE SYSTEM (K-FOLD CROSS VALIDATION WITH K = 10 (METRICS: PRECISION, RECALL, F1-SCORE) AND LEAVE-ONE-OUT CROSS-VALIDATION (LOOCV) FOR ACCURACY METRIC) FOR MULTIPLE MODELS AND EACH OF THE DATA TYPES FED TO THE SYSTEM: RAW DATA ONLY (R), QUATERNIONS DATA ONLY (Q), AND COMBINATION OF RAW AND QUATERNIONS DATA (R&Q). BEST PERFORMANCE IN 3/4 METRICS MARKED IN **BOLD**. (NOTE: FOR MULTI-CLASS CLASSIFIER WE USED THE WEIGHTED AVERAGE OF EACH METRIC ACROSS ALL 27 CLASSES, SEE TABLE I).

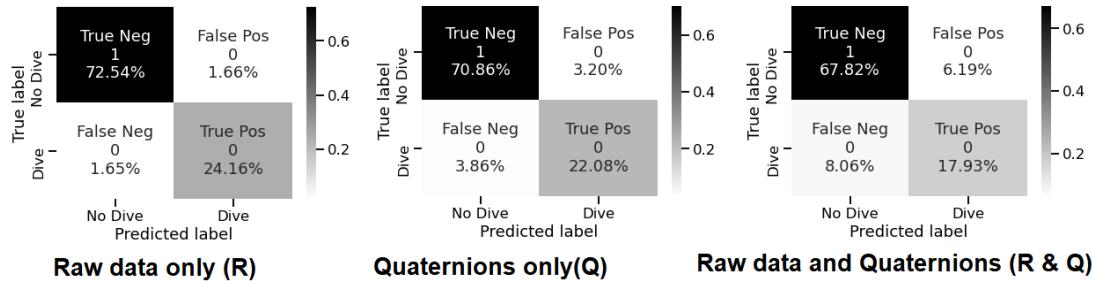


Fig. 4. Confusion matrices for the (binary) dive discrimination for each of the data types fed to the system: raw data only (R), quaternions data only (Q), and the combination of raw and quaternions data (R&Q).

LOOCV Accuracy).

- Rank experiments depending on performance.
- Perform statistical tests (i.e. a combination of omnibus ANOVA and posthoc pairwise T-test with a significance $p = 0.05$) and adjust ranking depending on significance.
- Evaluate best algorithms depending on ranking for subsets of relevant metrics (i.e. the metrics with significant difference).

The results of the significance analysis are depicted in Figure 5, where baseline is just a simple Gaussian Naive Bayes classifier.

Explainability Analysis

Interpretability is the degree to which a human can understand the cause of a machine learning decision. Moreover, an explanation relates the feature values of an instance to the model prediction in a humanly understandable way. In this section, we complement our performance evaluation of our end-to-end system with an analysis of the actual explanations of the predictions. SHAP uses the conditional expectation to estimate effects of the features in the final prediction, capturing also cases when a feature is correlated with another feature that actually has a stronger influence on the prediction.

For the explainability analysis, due to space constraints, we only look at the dive (binary) discrimination scenario in which our system was trained on the combination of raw and quaternions data. In all the other cases, the analysis follows the same principle and the supporting code is available for the interested reader on GITLAB⁶. The first aspect of the analysis looks at the combination of feature importance with feature effects on the prediction. Each point on the summary plot in Figure 6 is the impact of a feature and an instance (i.e. the feature itself). The position on the y-axis is determined by the feature and on the x-axis by the impact on the model prediction (i.e. Shapley value). The color in the left panel represents the value of the feature from low to high. Overlapping points provide a sense of the distribution of the Shapley values per feature. The features are ordered according to their importance, as shown in the right panel. Figure 6 provides insightful information about goalkeeper's kinematics. For instance, from the left panel, one can see that low magnetic field intensity on the z axis (i.e. mz is proportional with the absolute angle w.r.t z axis) has a strong contribution in the left/right dive discrimination. Moreover, high values of the first quaternion component (q1) and magnetic field on y axis (my) determine a low discrimination capability. From the right panel one can see that the combination of quaternion components (q1, q2, q3, q4) have a visible cumulative impact by changing the predicted dive type probability on average by 18 percentage points (i.e. $q1 + q2 + q3 + q4 = 0.035 + 0.025 + 0.06 + 0.06$).

Finally, we analyse how the model arrives at its decision by using a SHAP force plot. Basically, we decomposed a prediction into the sum of effects of each feature value. An example is depicted in Figure 7. Here we consider a random sample in the testing dataset of a certain goalkeeper and show how the model computed the predicted probability of

a dive. Important to note that the depicted effects describe the behavior of the model and are not necessarily causal in the real world.

IV. DISCUSSION

Our study investigates the extraction of accurate and explainable goalkeeper kinematics from a single motion sensor. Throughout our experiments we demonstrate the advantages of data-driven machine learning approaches in exploiting raw data for end-to-end motion understanding. We carry our discussion in the framework set by four recent and relevant studies. We focus on: the type of machine learning algorithm used, the sensory setup (e.g. type, number, and placement of sensors), and the explainability of the predictions, respectively. The choice of a machine learning algorithm is not always straightforward, requiring a good understanding of the domain the data comes from, the data peculiarities, and the actual task to solve. The flexibility of powerful learning algorithms can be sometimes hindered by not selecting the most informative sensory data for the task. In a first relevant study, reconstructing the full human body pose in real-time from 6 body-worn inertial sensors, [37] used recurrent deep neural networks to learn temporal pose priors, hence avoiding nonlinear optimization. The model proved to be robust for normal displacements and orientation, but failed in cases where the body is moving parallel to the floor (i.e. diving) as it did not fully exploit the acceleration information. Moreover, the system was sensitive to sensor number, placement, and noise characteristics. Our work used a single inertial sensor for a semantically simple task, goalkeeper dive discrimination and motion type segmentation (see Table I), which didn't require full-body pose, rather the identification of "signatures" of different types of motions as depicted in Figure 2. To achieve this, we used the XGBoost ensemble method which, due to its additive tree learning strategy, offers custom optimization objectives of weak learners, hence avoiding expensive nonlinear optimization. Our system is not sensitive to fast and noisy motions providing precise dive discrimination or motion segmentation by combining the raw and quaternions data, as shown in Table III.

Goalkeeper kinematics carry special features in terms of the timing of motions, explosive reactions, whole body motion, and a clear imbalance in the type of motions (i.e. the percentage of dives w.r.t. ball catches, ball throws, passes, and sprinting, respectively - see Figure 4. Using a series of online local dynamic models derived from the CMU motion database⁷ and trained on various motions (e.g. swings, punches), the work of [38] reconstructed human motion from 6 inertial sensors in a maximum a posteriori framework. Despite its flexibility in exploiting the acceleration information, the system was bound to lab experiments and only limited to motions that assume a vertical trunk position of the body (e.g. golf swing, boxing punch). The system needed additional calibration (performed with additional ultrasonic sensors) and need an offline optimization step (i.e. Levenberg-Marquadt) for

⁶<https://gitlab.com/akii-microlab/goalkeeper-motion-understanding-framework>

⁷<http://mocap.cs.cmu.edu/>

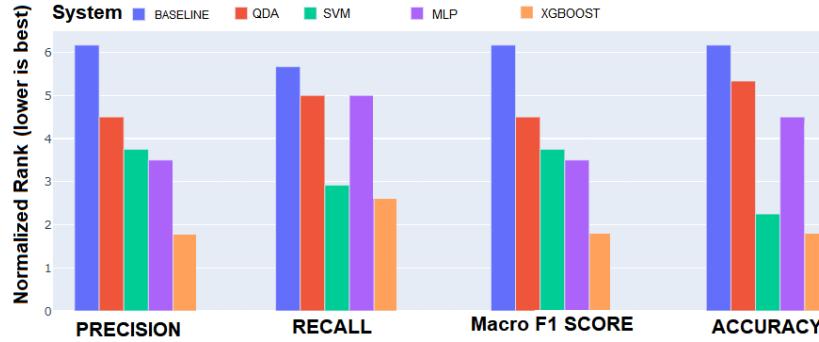


Fig. 5. Statistical ranking of the different approaches we evaluated.

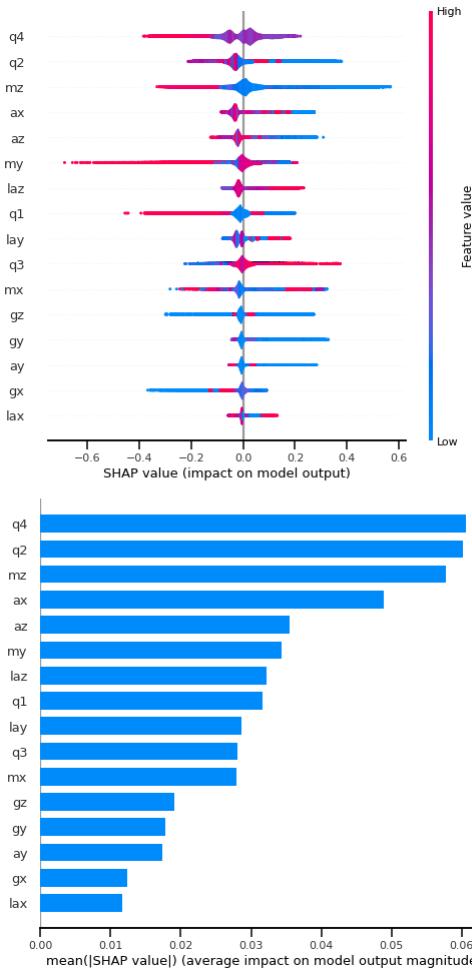


Fig. 6. SHAP explanation for the (binary) dive discrimination for the combination of raw and quaternions data (R&Q).

a marginal improvement over baseline (i.e. an efficient Online PCA), despite the large number of sensors.

In a study that combined camera data with inertial sensing in a Hierarchical Multivariate Hidden Markov Model, the system of [39] learned the structure of motion sequences from a single inertial sensor. The inertial sensor placement (i.e. ankle) only allowed for lower-body segmentation, which was then propagated through the Markov model to close the whole kinematics chain. Despite the well chosen model

to extract full body kinematics, the system used extensive feature engineering. Avoiding the full body kinematic chain and feature engineering, our work considered learning the sensory signatures of dives and other types of specific goalkeeper motions from informative raw data. The repertoire of motions in the [39] study was limited and conservative, without lateral accelerations and explosive jumps, making the system usable only in limited cases requiring reconstruction of simple motions. This limitation was imposed by the choice of the complex Hierarchical Multivariate Hidden Markov Model. We alleviate these constraints by using XGBoost that allowed the system to learn higher-order interactions between raw data, without expensive and bias-prone feature engineering, even from a single sensor.

Finally, in a goalkeeper dedicated solution, [14] developed a single sensor motion classification system. Similar to our work, the study only used a sensor mounted in the goalkeeper gloves. Opting for extensive data engineering, the authors developed complex "hand-crafted" features from raw data (i.e. various statistics and heuristics), subsequently ranked them, and fed them to a series of traditional machine learning algorithms (e.g. Support Vector Machines, Neural Networks). The study was mainly descriptive, evaluating how each algorithm performed on the pre-processed data, offering limited insights in the actual kinematics. The study didn't explain how the predictions map to actual insights in goalkeeper biomechanics, giving only a future outlook on the potential applicability of the system. In our study, we employed regression trees (i.e. XGBoost), which perform well for both classification and regression at scale and use an explainable learning approach. Such an approach can explain higher-order interactions between raw data streams, as shown in Figure 6. Additionally, each system prediction could account for the contribution each of the sensory data has in the final output, as shown in Figure 7. Such insights can be subsequently used to materialize complex analytics, for instance "the side of the post where most dives were performed" given the values of each sensory quantity (i.e. high contribution of mz, q2, my (i.e. y and z magnetometer values and second quaternion component) and low contribution of q4, laz and mx (i.e. forth quaternion component, z axis lateral acceleration and x axis magnetometer reading) characterising a dive - depicted in Figure 7).

Finally, the proposed end-to-end learning system is applicable to different types and number of sensors. Regardless

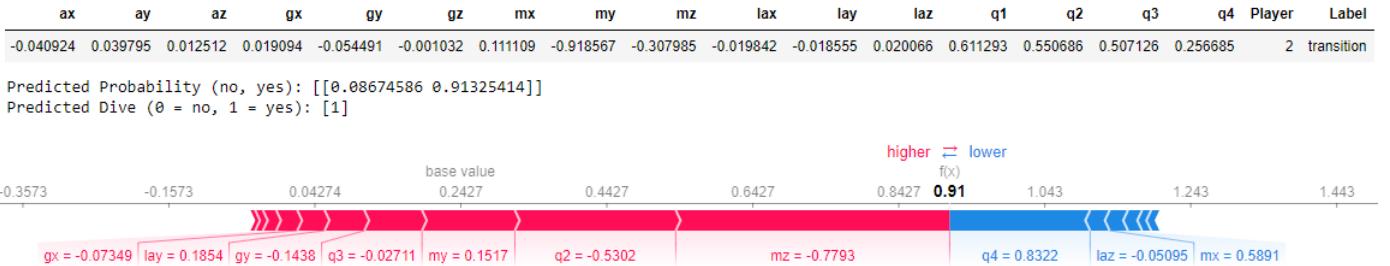


Fig. 7. Explanation of the model decision making for the (binary) dive classification for the combination of raw and quaternions data (R&Q).

of their placement, it can learn the interactions between available sensory quantities and the motion signatures they capture. This allows our system to exploit the motion regularities as perceived from all the different motion sensors (i.e. accelerometer, gyroscope, magnetometer) and synergistically combine them without bias-prone feature engineering. From the technical point of view of the implementation, the choice of XGBoost as machine learning algorithm is motivated by the need to deploy the system as standalone data acquisition and processing system. The machine learning algorithm can be used for inference in resource-limited devices such as the Goalplay embedded glove system using dedicated tools [51].

V. CONCLUSIONS

Despite the availability of both sensing hardware and processing algorithms, sport applications are still lacking accessible solutions which can operate on the field. Either confined to controlled lab environments or expensive on-field sensing infrastructures, the kinematic assessment of the athletes is still an open problem. In our study, we tackle this problem and introduce a methodology to process information from a single motion sensor module to successfully extract accurate and explainable motion understanding in the case of soccer goalkeepers. In order to exploit the richness and unique parameters of soccer goalkeepers motion, we employed a machine learning algorithm which was able to discriminate dives or other types of specific motions directly from raw sensory data. We further evaluate multiple algorithms through statistical significance tests to be able to find an optimal configuration. Using real-world data from 7 goalkeepers, we trained our end-to-end machine learning system to classify motion and at the same time explain how each sensed motion component contributes to describing a specific goalkeeper action. The potential for real-world deployment is high. The system offers a cheap and unobtrusive system for accurate and explainable decision making based solely on raw data with the potential for providing personalized assessment. This system is meant to perform an assessment of the training session which is typically designed by a trainer. Basically, the trainer makes the exercise plan and the system provides a quantitative and qualitative assessment of the training outcome.

VI. ACKNOWLEDGEMENT(S)

We would like to thank the goalkeepers, goalkeeper trainers, and the Goalplay GmbH - Co. KG software team for collecting, annotating, and curating the datasets.

VII. FUNDING

The work is part of PERSEUS Project (Platform for Enhanced virtual Reality in Sport Exercise Understanding and Simulation, No. ZF4017410SS9) funded by the Central Innovation Program for Small and Medium-sized Enterprises (ZIM) from the German Federal Ministry for Economic Affairs and Energy (BMWi).

VIII. NOTES ON CONTRIBUTIONS

GL designed the research, implemented the data analysis algorithms, wrote the first draft of the manuscript. CP contributed to goalkeeper exercise design, performed data acquisition, and developed data annotation software. TG designed data analysis algorithms, wrote the first draft of the manuscript. CA designed the research, data analysis, and wrote the first draft of the manuscript. All authors wrote and approved the manuscript.

REFERENCES

- [1] Y. Wang, Q. Ji, and C. Zhou, "Effect of prior cues on action anticipation in soccer goalkeepers," *Psychology of Sport and Exercise*, vol. 43, pp. 137–143, 2019.
- [2] F. W. Otte, S.-K. Millar, and S. Klatt, "How does the modern football goalkeeper train?—an exploration of expert goalkeeper coaches' skill training approaches," *Journal of sports sciences*, vol. 38, no. 11-12, pp. 1465–1473, 2020.
- [3] M. Knoop, J. Fernandez-Fernandez, and A. Ferrauti, "Evaluation of a specific reaction and action speed test for the soccer goalkeeper," *The Journal of Strength & Conditioning Research*, vol. 27, no. 8, pp. 2141–2148, 2013.
- [4] J. J. Malone, A. Jaspers, W. Helsen, B. Merks, W. G. Frencken, and M. S. Brink, "Seasonal training load and wellness monitoring in a professional soccer goalkeeper," *International Journal of Sports Physiology and Performance*, vol. 13, no. 5, pp. 672–675, 2018.
- [5] J. Rodríguez-Arce, L. I. Flores-Núñez, O. Portillo-Rodríguez, and S. E. Hernández-López, "Assessing the performance of soccer goalkeepers based on their cognitive and motor skills," *International Journal of Performance Analysis in Sport*, vol. 19, no. 5, pp. 655–671, 2019.
- [6] K.-U. Schmitt, M. Schlüttler, and P. Boesiger, "Biomechanical loading of the hip during side jumps by soccer goalkeepers," *Journal of sports sciences*, vol. 28, no. 1, pp. 53–59, 2010.
- [7] G. J. Diaz, B. R. Fajen, and F. Phillips, "Anticipation from biological motion: the goalkeeper problem," *Journal of experimental psychology: human perception and performance*, vol. 38, no. 4, p. 848, 2012.
- [8] J. A. Navia, J. Van der Kamp, and L. M. Ruiz, "On the use of situational and body information in goalkeeper actions during a soccer penalty kick," *International Journal of Sport Psychology*, vol. 44, no. 3, pp. 234–251, 2013.
- [9] O. M. Labib and M. S. Mahmoud, "Kinetic analysis for performance methods of long kick skill as a criterion to select soccer goalkeeper," *Journal of Applied Sports Science*, vol. 7, no. 1, 2017.
- [10] S. Suzuki, H. Togari, M. Isokawa, J. Ohashi, and T. Ohgushi, "Analysis of the goalkeeper's diving motion," in *Science and Football: Proceedings of the First World Congress of Science and Football Liverpool*, 2011.

- [11] H. Sørensen, M. Thomassen, and M. Zacho, "Biomechanical profile of soccer goalkeepers," in *ISBS-Conference Proceedings Archive*, 2001.
- [12] N. Numazu and N. Fujii, "Biomechanical analysis of saving motion for soccer goalkeepers focused on the function of lower extremities," in *ISBS-Conference Proceedings Archive*, 2015.
- [13] D. G. Kerwin and K. Bray, "Measuring and modelling the goalkeeper's diving envelope in a penalty kick," in *The Engineering of sport 6*. Springer, 2006, pp. 321–326.
- [14] J. Haladjian, D. Schlabbers, S. Taheri, M. Tharr, and B. Bruegge, "Sensor-based detection and classification of soccer goalkeeper training exercises," *ACM Transactions on Internet of Things*, vol. 1, no. 2, pp. 1–20, 2020.
- [15] H. Liu, M. A. Gómez, and C. Lago-Peña, "Match performance profiles of goalkeepers of elite football teams," *International Journal of Sports Science & Coaching*, vol. 10, no. 4, pp. 669–682, 2015.
- [16] VICON, "Visual motion tracker," 2020-11-10. [Online]. Available: <https://www.vicon.com/applications/life-sciences/>
- [17] ———, "Trident imu motion tracker," 2020-11-10. [Online]. Available: <https://www.vicon.com/hardware/blue-trident/>
- [18] XSENS, "Dot imu motion tracker," 2020-11-10. [Online]. Available: <https://www.xsens.com/xsens-dot>
- [19] STT-Systems, "isen motion tracker," 2020-11-10. [Online]. Available: <https://www.stt-systems.com/motion-analysis/inertial-motion-capture/isen/?cn-reloaded=1>
- [20] Zepp-Labs, "Play soccer tracker," 2020-11-10. [Online]. Available: <http://www.zeplabs.com/en-us/soccer>
- [21] R. Rebelo-Gonçalves, A. J. Figueiredo, M. J. Coelho-e Silva, and A. Tessitore, "Assessment of technical skills in young soccer goalkeepers: Reliability and validity of two goalkeeper-specific tests," *Journal of sports science & medicine*, vol. 15, no. 3, p. 516, 2016.
- [22] W. Spratford, R. Mellifont, and B. Burkett, "The influence of dive direction on the movement characteristics for elite football goalkeepers," *Sports Biomechanics*, vol. 8, no. 3, pp. 235–244, 2009.
- [23] N. Numazu, N. Fujii, M. Nakayam, and M. Koido, "Game performance analysis of soccer goalkeepers comparison between saving motion and other motions," in *ISBS-Conference Proceedings Archive*, 2016.
- [24] M. Rana and V. Mittal, "Wearable sensors for real-time kinematics analysis in sports: A review," *IEEE Sensors Journal*, 2020.
- [25] A. Konrad and M. Tilp, "Muscle and tendon tissue properties of competitive soccer goalkeepers and midfielders," *German Journal of Exercise and Sport Research*, vol. 48, no. 2, pp. 245–251, 2018.
- [26] S. S. Tabrizi, S. Pashazadeh, and V. Javani, "Comparative study of table tennis forehand strokes classification using deep learning and svm," *IEEE Sensors Journal*, vol. 20, no. 22, pp. 13 552–13 561, 2020.
- [27] C. Lago-Peña, J. Lago-Ballesteros, and E. Rey, "Differences in performance indicators between winning and losing teams in the uefa champions league," *Journal of human kinetics*, vol. 27, no. 2011, pp. 135–146, 2011.
- [28] D. Schuldhaus, C. Zwick, H. Koerger, E. Dorschky, R. Kirk, and B. M. Eskofier, "Inertial sensor-based approach for shot/pass classification during a soccer match," in *KDD Workshop on Large-Scale Sports Analytics*, 2015, pp. 1–4.
- [29] H. K. Stensland, V. R. Gaddam, M. Tennøe, E. Helgedagsrud, M. Næss, H. K. Alstad, A. Mortensen, R. Langseth, S. Ljødal, Ø. Landsverk *et al.*, "Bagadus: An integrated real-time system for soccer analytics," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 10, no. 1s, pp. 1–21, 2014.
- [30] H. Xu, J. Liu, H. Hu, and Y. Zhang, "Wearable sensor-based human activity recognition method with multi-features extracted from hilbert-huang transform," *Sensors*, vol. 16, no. 12, p. 2048, 2016.
- [31] B. Zhou, H. Koerger, M. Wirth, C. Zwick, C. Martindale, H. Cruz, B. Eskofier, and P. Lukowicz, "Smart soccer shoe: monitoring football interaction with shoe integrated textile pressure sensor matrix," in *Proceedings of the 2016 ACM International Symposium on Wearable Computers*, 2016, pp. 64–71.
- [32] L. A. Schwarz, D. Mateus, and N. Navab, "Recognizing multiple human activities and tracking full-body pose in unconstrained environments," *Pattern Recognition*, vol. 45, no. 1, pp. 11–23, 2012.
- [33] A. White, S. P. Hills, C. B. Cooke, T. Batten, L. P. Kilduff, C. J. Cook, C. Roberts, and M. Russell, "Match-play and performance test responses of soccer goalkeepers: a review of current literature," *Sports Medicine*, vol. 48, no. 11, pp. 2497–2516, 2018.
- [34] kinexon, "Real-time data collection and analysis for football performance," 2020-11-10. [Online]. Available: <https://kinexon.com/de/loesungen/fussball>
- [35] W. Kim and M. Kim, "On-line detection and segmentation of sports motions using a wearable sensor," *Sensors*, vol. 18, no. 3, p. 913, 2018.
- [36] E. Preatoni, S. Nodari, and N. Lopomo, "Supervised machine learning applied to wearable sensor data can accurately classify functional fitness exercises within a continuous workout," *Frontiers in Bioengineering and Biotechnology*, 2020.
- [37] Y. Huang, M. Kaufmann, E. Aksan, M. J. Black, O. Hilliges, and G. Pons-Moll, "Deep inertial poser: learning to reconstruct human pose from sparse inertial measurements in real time," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 6, pp. 1–15, 2018.
- [38] H. Liu, X. Wei, J. Chai, I. Ha, and T. Rhee, "Realtime human motion control with a small number of inertial sensors," in *Symposium on interactive 3D graphics and games*, 2011, pp. 133–140.
- [39] C. Mousas, "Full-body locomotion reconstruction of virtual characters using a single inertial measurement unit," *Sensors*, vol. 17, no. 11, p. 2589, 2017.
- [40] L. A. Schwarz, D. Mateus, and N. Navab, "Discriminative human full-body pose estimation from wearable inertial sensor data," in *3D physiological human workshop*. Springer, 2009, pp. 159–172.
- [41] I. Gotlibovych, S. Crawford, D. Goyal, J. Liu, Y. Kerem, D. Benaron, D. Yilmaz, G. Marcus, and Y. Li, "End-to-end deep learning from raw sensor data: Atrial fibrillation detection using wearables," *arXiv arXiv:1807.10707*, 2018.
- [42] R. Ibrahim, I. Kingma, V. A. de Boode, G. S. Faber, and J. H. van Dieën, "Kinematic and kinetic analysis of the goalkeeper's diving save in football," *Journal of sports sciences*, vol. 37, no. 3, pp. 313–321, 2019.
- [43] K.-U. Schmitt, M. Nusser, and P. Boesiger, "Hip injuries in professional and amateur soccer goalkeepers," *Sportverletzung Sportschaden: Organ der Gesellschaft für Orthopädisch-Traumatologische Sportmedizin*, vol. 22, no. 3, pp. 159–163, 2008.
- [44] C. Prodaniuc, "Goalplay wearable data," Mar. 2021. [Online]. Available: <https://doi.org/10.5281/zenodo.4629325>
- [45] A. Arac, "Machine learning for 3d kinematic analysis of movements in neurorehabilitation," *Current Neurology and Neuroscience Reports*, vol. 20, no. 8, pp. 1–6, 2020.
- [46] R. Villegas, J. Yang, D. Ceylan, and H. Lee, "Neural kinematic networks for unsupervised motion retargetting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8639–8648.
- [47] J. Figueiredo, S. P. Carvalho, D. Gonçalve, J. C. Moreno, and C. P. Santos, "Daily locomotion recognition and prediction: A kinematic data-based machine learning approach," *IEEE Access*, vol. 8, pp. 33 250–33 262, 2020.
- [48] D. K. Dennis, D. Acar, V. Mandikal, V. Sadasivan, H. Simhadri, V. Saligrama, and P. Jain, "Shallow rnns: a method for accurate time-series classification on tiny devices," 2019.
- [49] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in *EANN*, vol. 3, 2013, p. 3.
- [50] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 785–794.
- [51] H. Cho and M. Li, "Treelite: toolbox for decision tree deployment," 2018.
- [52] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in neural information processing systems*, 2017, pp. 4765–4774.
- [53] S. M. Lundberg, G. G. Erion, and S.-I. Lee, "Consistent individualized feature attribution for tree ensembles," *arXiv preprint arXiv:1802.03888*, 2018.
- [54] H. He, Y. Bai, E. A. Garcia, and S. Li, "Adasyn: Adaptive synthetic sampling approach for imbalanced learning," in *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*. IEEE, 2008, pp. 1322–1328.