VILNIUS UNIVERSITY
FACULTY OF MATHEMATICS AND INFORMATICS
DATA SCIENCE STUDY PROGRAMME

Master's thesis

Analysis of action recognition using skeleton data

Veiksmų atpažinimo analizė naudojant skeleto duomenis

Hermanas Vaitiekus

Supervisor : assoc. prof. Linas Petkevičius
Reviewer : prof. Povilas Treigys

Vilnius
2025

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Linas Petkevičius, for his guidance, support, and insightful feedback throughout this research. From the initial idea to the very end, his expertise, patience, and dedication were invaluable in shaping and completing this thesis. His continuous involvement greatly contributed to the success of this work.

I am also thankful to the Faculty of Mathematics and Informatics, Vilnius University for providing an inspiring academic environment. Despite taking breaks in my studies, the faculty's support and flexibility enabled me to successfully continue and complete my work. The knowledge and skills I gained throughout my studies have been essential in carrying out this research and bring it to completion.

A special thanks to the CrossFit Laisvė community, whose environment and inspiration sparked the idea for this master thesis. The supportive atmosphere and dedication to improving together provided the motivation to undertake this project. I would also like to express my gratitude to the HWPO community for instilling the "Hard Work Pays Off" mentality, which encouraged me to push through challenging days and stay committed to excellence.

Finally, I would like to thank my girlfriend, family, and friends for their unwavering support and understanding during this academic journey.

Summary

This thesis focused on one-class classification of actions video records. The thesis covers developing a framework for analyzing snatch lifts in CrossFit, utilizing skeleton data and Dynamic Time Warping (DTW) to assess the quality of the lift. The objective was to provide a data driven evaluation of the movement, complementing traditional coaching methods. The study began with skeleton data extraction, followed by data normalization techniques, including scaling, centering, rotation, and smoothing. A golden standard was created and the performance of other lifts was compared against this standard using DTW. Outliers were detected to identify lifts with significant deviations from optimal form. The results demonstrated that the proposed framework identifies deviations from the golden standard, offering a promising tool. This study contributes to the growing body of research in sports performance analysis, particularly in the context of CrossFit.

Keywords: Video classification, One-Class clasification, Snatch lift, CrossFit, Movement analysis, Skeleton data, Dynamic Time Warping (DTW), Outlier detection

Santrauka

Šis darbas buvo orientuotas į vienos klasės veiksmų vaizdo įrašų klasifikaciją. Baigiamasis darbas apima proceso, skirto rovimo technikos analyzei, kūrimą. Jame pasitelkiama skeleto struktūros duomenys bei dinaminis laiko skalės iškraipymo (DTW) metodas. Darbo tikslas buvo pateikti duomenimis grįstą judesių vertinimą, papildant tradicines treniravimo metodikas. Buvo sukurtas auksinis standartas - etaloninis kėlimas, su kuriuo buvo lyginami kiti kėlimai. Darbas buvo pradėtas skeleto duomenų išgavimu, po kurio sekė duomenų normalizavimas, įskaitant mastelio keitimą, centravimą, rotaciją ir glotninimą. Siekiant nustatyti, ar kėlimų technika nukrypsta nuo optimalios atlikimo techikos, buvo identifikuotos išskirtys. Resultatai parodė, kad siūlomas procesas leidžia identifikuoti nukrypimus nuo auksinio standarto. Šiuo darbu prisidedama prie didėjančio sporto rezultatų analyzės tyrimų skaičiaus, ypač krosfito kontekste.

Raktiniai žodžiai: Vaizdo įrašų klasifikavimas, Vienos klasės klasifikavimas, Rovimo veiksmas, Krosfitas, Judesių analyzė, Skeleto duomenys, Dinaminis laiko skalės iškraipymas (DTW), Išskirčių identifikavimas

Contents

List of Figures

List of Tables

Introduction

Background and context

CrossFit has seen a rapid increase in popularity, with millions of participants worldwide embracing its high-intensity and varied training methodology. Among the numerous exercises performed in Crossfit, the snatch lift stands out for its technical complexity [8] and demand for precise coordination. It is frequently included in Crossfit competitions and training programs, making it a critical movement for athletes to master.

Despite its significance, the snatch lift is challenging to execute correctly. Errors in technique not only reduce performance, but also increase the risk of injury [7]. Traditional coaching methods often rely on subjective observation, which can be limited in providing detailed feedback.

This has led to an interest in the application of technology for movement analysis. Skeleton data, which represents human motion through key-points, provides an efficient and precise method for quantifying and analyzing complex movements. By leveraging such data, this study aims to provide a framework for evaluating the quality of snatch lifts, identifying deviations from optimal execution, and offering actionable insights for athletes and coaches alike.

Problem statement

Video classification of complex athletic movements, such as the snatch lift, presents a significant challenge in sports performance evaluation. The analysis of complex athletic movements such as the snatch lift poses significant challenges in sports performance evaluation. This movement, central to Olympic weightlifting and Crossfit, requires precise coordination and technical skill. Currently, the assessment of snatch lifts relies heavily on the subjective judgment of coaches, which can vary and lead to inconsistencies, especially for athletes training without direct supervision. Without an objective and standardized approach, athletes may develop flawed techniques, potentially resulting in performance decline and increased risk of injury.

There is a critical need for reliable, data-driven tools to assess the technical quality of snatch lifts, particularly for athletes who do not have access to immediate coach feedback.

Aim of the thesis

The aim of the thesis is propose and develop one-class correct snatch lift classification method for video records.

Research Objectives

The primary objective of this research is to develop a methodology for analyzing snatch lifts using skeleton data. Specifically, the study aims to:

1. Research and identify the most suitable skeleton recognition method for analyzing the snatch lift;

2. Process video data through the selected skeleton recognition model and normalize it for consistent scale, alignment, and smoothness;

3. Create a golden standard for technically correct snatch lifts using average aggregation methods;

4. Compare lifts against the golden standard using Dynamic Time Warping;

5. Detect and analyze outliers by identifying significant deviations from the golden standard.

Significance of the Study

The significance of this study lies in its contribution to improving the accuracy and objectivity of motion analysis in sports, specifically in evaluating snatch lift. Using skeleton-based action recognition and dynamic time warping, this research offers a novel framework to objectively assess the technical quality of snatch lifts, addressing the limitations of current methods, which are often subjective and inconsistent.

The proposed framework provides athlete with an innovative tool to independently evaluate their lifts, ensuring more consistent technique over time, even in the absence of real-time supervision. Ultimately, this approach helps to minimize the risk of injury and enhance performance by offering a standardized method for comparing lifts against a 'golden standard' of correct movement.

Beyond individual performance enhancement, the findings of this research have broader implications for sports analytics. The methods developed here can be applied to other complex movements in various sports, advancing the field of action recognition and offering new opportunities for research in motion analysis. This work also contributes to the growing body of knowledge in automated performance evaluation, potentially influencing future developments in sports technology and biomechanics.

Overview of the Thesis

This thesis is structured into five chapters, each addressing a key component of the research process, from theory and methodology to analysis and conclusions.

- *Chapter 1: Introduction* - The introductory chapter provides an overview of the research problem, the significance of the study, and the objectives to be achived. It sets the stage for understanding the challenges in analyzing snatch lifts and introduces the framework developed in this research.

- *Chapter 2: Literature Review* - This chapter reviews existing research on motion analysis techniques and skeleton-based models in sports performance evaluation. The literature review establishes the theoretical foundation for the proposed research.

- *Chapter 3: Methodology* - The methodology chapter outlines the research design, including the data collection process, the skeleton recognition models used, and the techniques for normalizing and comparing snatch lifts using dynamic time warping. This chapter details how the golden standard is created and how lifts are analyzed against it.

- *Chapter 4: Experiment* - The experiment chapter demonstrates the practical application of the methodology, detailing each step performed in a Google Colab environment. It highlights the implementation of skeleton data extraction, normalization, golden standard creation and DTW-based comparison.

- *Chapter 5: Results* - The results chapter presents the findings from applying the proposed framework to the snatch lift data, including comparisons of multiple lifts and the identification of outliers.

- *Chapter 6: Conclusions* - The final chapter reflects on the achievements of the study, summarizing key findings and their relevance. It discusses the study's limitations and provides recommendations for future research, along with practical applications of the developed framework.

1 Literature Review

In recent years, there has been significant progress in the development of techniques for human pose estimation, which are critical for evaluating comlpex physical activities such as snatch lift in weightlifting. This chapter reviews the existing literature on key technologies relevant to this study, including pose estimation models and time series analysis techniques such as Dynamic Time Warping (DTW).

1.1 Human Pose Estimation

Human pose estimation is a vital task in motion analysis, as it enables the tracking of key points or joints on the human body in order to capture and evaluate human motion. A variety of methods and models have been developed for both 2D and 3D pose estimation, each offering distinct advantages and challenges.

One of the models in this domain is *UniPose* [1], a unified framework designed to perform human pose estimation in both single image and videos. This model focuses on extracting accurate 2D and 3D human pose information, making it suitable for complex real-time applications, including sports performance analysis. The model's ability to work with both images and videos is especially useful in this research, where high-quality video data is used to extract pose information for analyzing snatch lifts. *Unipose+* [3] enhances the original UniPose model by introducing multi-scale feature extraction, refined heat-map modulation, support for 3D pose estimation, compatibility with multiple backbone networks, and improved overall performances. These models provide a strong foundation for extracting skeleton data in human movement analysis, making them particularly relevan for evaluating the technical execution of weightlifting techniques like the snatch lift.

Another notable approach for human pose estimation is *OmniPose* [2], a multi-scale framework designed for multi-person pose estimation. While this model is primarily intended for scenarios involving multiple individuals, its ability to handle multiple scales and occlusions allows for better performance in complex environments where various body parts may be partially obscured. For this study, the insights provided by OmniPose are useful in understanding how different pose models can be adapted to various video environments, ensuring accurate skeleton extraction even under challenging conditions.

In addition to these frameworks, recent advancements in pose estimation have incorporated deep learning techniques and multi-source datasets. For instance, *Sarandi, A. Hermans and B. Leibe*[9] introduced a geometry-aware autoencoder that bridges multiple skeleton formats for 3D human pose estimation. This model learns from a large number of datasets to accurately predict 3D human poses, even in scenarios with challengig data. This model's ability to handle different skeleton formats is particularly relevant for this research. This model, which serves as the backbone of the skeleton extraction process in this thesis, was crucial in ensuring that the snatch lift data could be accurately analyzed from different video sources.

1.2 Dynamic Time Warping (DTW) for Motion Analysis

Dynamic Time Warping (DTW) is a time series analysis technique that measures the similarity between two sequences by aligning them optimally, even if they are of different lengths or occur at different speeds. DTW has been widely used in various applications, from speech recognition to motion analysis, due to its robustness in handling time-series data with varying temporal properties.

Giorgino[6] provides a comprehensive overview of DTW and its implementation in R, with a particular focus on the dtw package, which is commonly used for visualizing and computing DTW alignments. The ability to compute DTW distances between time series sequences makes it an ideal technique for comparing the trajectories of joint movements in weightlifting, where the execution of the snatch lift can vary across lifters. By using DTW, this research can compare the skeletal trajectories of key joints during the snatch lift and assess deviations from an optimal execution. The normalized distance computed by DTW enables the identification of lifts that deviate significantly from the golden standard, offering a quantitative basis for evaluating lift quality.The DTW process consists of the following steps:

1. *Input Sequences:*
   Two sequences of joint coordinates serve as input:
   - $Q = [q_1, q_2, \ldots, q_n]$: The query sequence, representing the skeleton data of the lift under evaluation.
   - $R = [r_1, r_2, \ldots, r_m]$: The reference sequence, representing the skeleton data of the "golden standard" lift.

2. *Cost Matrix Calculation:*
   A cost matrix C of dimensions $n \times m$ is constructed, where each entry $C(i, j)$ represents the distance between the i-th point of sequence Q and the j-th point of sequence R. Here, the Euclidean distance between joint coordinates is used:

   $$C(i, j) = \|q_i - r_j\| \tag{1}$$

   where q_i and r_j are vectors representing joint coordinates (e.g., 2D or 3D).

3. *Accumulated Cost Matrix:*
   An accumulated cost matrix D is computed iteratively to store the minimum cost required to align the first i-th points of Q with the first j-th points of R. The recurrence relation is:

   $$D(i, j) = C(i, j) + \min \begin{cases} D(i-1, j) & \text{(insertion)} \\ D(i, j-1) & \text{(deletion)} \\ D(i-1, j-1) & \text{(match)} \end{cases} \tag{2}$$

with the following initial conditions:

$$D(0,0) = C(0,0), \quad D(0,j) = \sum_{k=1}^{j} C(0,k), \quad D(i,0) = \sum_{k=1}^{i} C(k,0) \tag{3}$$

4. *Optimal Warping Path:*
   The optimal warping path W is determined by tracing back from $D(n,m)$ (top-right corner) to $D(0,0)$ (bottom-left corner), following the path of minimal cumulative cost.

5. *Normalized Distance:*
   To account for differences in sequence lengths, the DTW distance is normalized by the number of steps in the warping path $|W|$:

$$\text{Normalized Distance} = \frac{\text{DTW Distance}}{|W|} \tag{4}$$

   This normalization ensures comparability between sequences of varying lengths.

6. *Visualization and Analysis:*
   The alignment results are visualized to verify that similar phases of the lift are matched correctly. This step is critical for assessing the practical utility of the DTW algorithm in evaluating motion. This step is illustrated in the Figure 2.

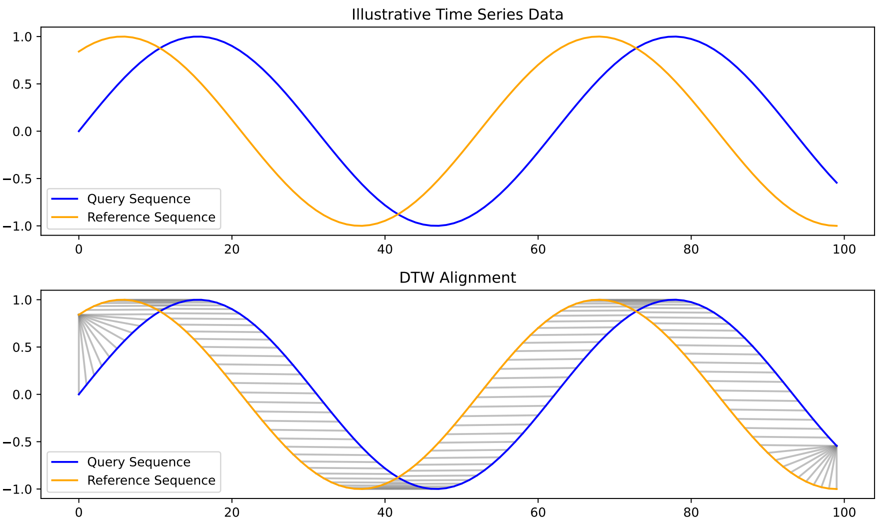

Figure 1. *Example of the DTW Alignment for the Illustrative Time Series Data*

1.3 Normalization Techniques

In this study, normalization of the skeleton data is performed using Bezier curves and rotation matrices to standardize and align the data for comparison. Bezier curves are commonly employed in motion analysis for their ability to smooth and interpolate between key points, ensuring that motion trajectories are continuous and free from abrupt fluctuations. A

Bezier curves are parametric curves based on n-th order Bernstein polynomial, where the time argument t belongs to the interval [0;1] [4].

$$B(t) = \sum_{i=0}^{n} \binom{n}{i} t^i (1-t)^{n-i}, \quad 0 \leq t \leq 1 \tag{5}$$

or

$$B_n(t) = \sum_{i=0}^{n} B_i^n(t) \tag{6}$$

Figure 2. illustrates an example with control points $P_0, P_1, ...P_n$, where $n = 3$. The product of these points with the corresponding Bernstein polynomial terms represents the value of the Bezier cureve for the parameter t:

$$B(t) = \sum_{i=0}^{n} \binom{n}{i} t^i (1-t)^{n-i} P_i, \quad 0 \leq t \leq 1 \tag{7}$$

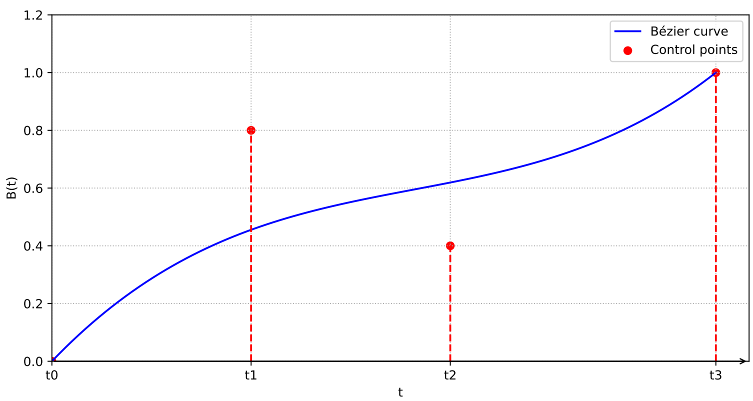

Figure 2. *Example of the Bezier curve for the time moments* $t_0 = 0, t_1 = 0.3, t_2 = 0.6, t_3 = 1$

These curves have been widely used in fields such as animation and biomechanics to refine movement data [5].

Rotation matrices, which are fundamental in spatial transformations, are utilized here to align the joint data in a consistent orientation. By applying rotation matrices, the variability introduced by different angles of capture is minimized, allowing for more accurate comparisons between lifts. This technique has been extensively used in computer vision and sports biomechanics [10].

2 Methodology

2.1 Overview of the Methodology

This chapter present the methodological framework adopted to analyze snatch lifts using skeleton-based motion data. The approach is designed to address the research objectives by ensuring accurate data collection, preprocessing and analysis. Specifically, the methodology focuses on extracting skeleton data from video recordings, normalizing and smoothing this data for consistency, and comparing individual lifts agains a golden standard using advance techniques.

The process begins with sourcing video data and using a pre-trained skeleton recognition model, which incorporates a geometry-aware autoencoder, to extract skeleton representations. The data is then normalized in terms of scale, alignment and smoothness to prepare it for analysis. Successful lifts are aggregated to create a golden standard representing optimal technique. Using Dynamic Time Warping, individual lifts are compared to this standard, and deviations are analyzed to identify outliers.

This methodological framework was chosen for its ability to combine advance motion recognition with systematic analysis, providing a robust and objective tool for assessing snatch lift quality. The following sections provide a detailed account for each step in this process.

2.2 Data collection and Preprocessing

2.2.1 Data source and collections

The dataset used in this study consists of 26 snatch lift videos sourced from the YouTube channel "Jėgos portas tiesiogiai", which provides live streams of the Lietuvos sunkiosios atletikos čempionatas (Lithuanian Weightlifting Championship). These videos feature athletes performing Olympic lifts under competetive conditions. Additionally, 20 videos were extracted from full competition videos on the "Weightlifting House" Youtube channel, specifically from the following events:

- Men's + 109 | World Weightlifting Championships 2023
- Women's -71kg Snatch | World Weightlifting Championships 2023
- Men's - 109 Snatch | World Weightlifting Championships 2023
- Women's +87 Snatch | World Weightlifting Championships 2023

Figure 3. *Samples of snatch lifts*

The dataset as showed in Figure 3. includes lifts from athletes of varying skill levels, ranging from local competitive lifters to top-tier international competitors, providing a diverse representation of techniques. All videos were downloaded from YouTube and trimmed to isolate the snatch lifts using the video editing software Shortcut. The trimmed video files were then uploaded to Google Drive, where they were made publicly accessible via specific links to ensure ease of access and sharing for analysis purposes. Majority of the videos were recorded in high definition (1080) at 25 fps to ensure visual clarity and consistency across the dataset.

2.2.2 Skeleton data extraction

To extract the skeleton data for analysis, a skeleton recognition model was applied to the preprocessed video footage. The model used in this study was based on a geometry-aware autoencoder as described in the article [9]. The model, which was downloaded from Tensorflow Hub, was applied to each frame of the video to detect human poses. Specifically, the "mpi_inf_3dhp_17" skeleton format was used, which tracks 17 key points, including joints such as the shoulders, elbows, hips and knees. The model outputs the 3D coordinates (x,y,z) of each keypoint, providing a comprehensive representation of the human body's movements. The choice of this model was based on its high accuracy in 3D pose estimation and its training on multiple validation datasets, which provided access to various skeleton formats, enhancing its suitability for analyzing weightlifting movements.

The extracted skeleton data utilized the "mpi_inf_3dhp_17" skeleton format, as showed in Figure 4., which derived from the MPI-INF-3DHP dataset, captures a total of 17 key points corresponding to major human joints. The MPI-INF-3DHP dataset itself was developed in a controlled lab environment using multiple cameras, offering high-quality annotations and comprehensive coverage of human motion. Additionally, the dataset includes both indoor and outdoor scenarios with varying degrees of occlusion, ensuring the model's adaptability to different conditions.

Figure 4. *Extracted skeletons for sample lifts. Videos recorded from front or angle of athlete*

While this skeleton format provides sufficient detail for analyzing complex movements like the snatch lift, it balances the inclusion of key points with the noise introduced during estimation. Incorporating more key points can lead to higher levels of uncertainty, particularly in challenging frames, making the "mpi_inf_3dhp_17" format an optimal choice for this study.

The skeleton extraction process was implemented using a python-based pipeline, leveraging a pose estimation model to process video frames iteratively. Each video was loaded and split into individual frames using OpenCV. For each frame, the model detected the positions of 17 key joints according to the "mpi_inf_3dhp_17" skeleton format. The detected joint coordinates - represented as (x,y,z) values - were saved frame by frame into JSON file for subsequent analysis. To ensure a consistent output, the pipeline also annotated frames. This lead to original data vectors for each video $i \in \{1, .., N\}$ and frame $f \in \{1, .., F_i\}$:

$$\vec{X}_{orig,f,i} = (x_{1,f,i}, y_{1,f,i}, z_{1,f,i}, .., x_{17,f,i}, y_{17,f,i}, z_{17,f,i})^T \tag{8}$$

The pipeline automatically handled multiple video samples by iterating through each video file, generating a dictionary of results where each frame was associated with its detected key points. These results were serialized and stored, forming a structured dataset representing the sequential motion of each lift. This automated approach facilitated efficient preprocessing of the video data, ensuring that the extracted skeleton data was well-organized and ready for further analysis.

2.2.3 Data normalization steps

The normalization process for the extracted skeleton data involved several key steps to ensure consistency and accuracy across all video. One of the first challenges encountered was the presence of missing data points in certain frames, which is common in pose estimation tasks, especially when joints are occluded or not detected clearly. To address this, the missing joint positions were interpolated using the previous frame's data, ensuring smooth continuity of the motion. This approach helped to avoid gaps in the movement data, which could negatively impact the analysis.

After addressing the missing data through interpolation, the next step in the normalization process was to scale and center the skeleton data to a consistent range. To achieve this, a custom function was implemented to first center the data by subtracting the mean of each joint's coordinates (x, y, z). This ensured that the data was aligned relative to the center of the body, making it independent of the specific position of the person in the video frame.

Next, the data was scaled to the range of [-1, 1]. This was done by calculating the minimum and maximum values for each joint's coordinates and dividing the centered data by the range. This scaling process ensured that all the movements across different frames and videos were normalized, allowing for consistent comparisons between them.

The scaling and centering process was applied to all frames in the dataset. The loop iterated over each video sample, extracting the 3D poses and applying the scaling and centering function.

To ensure consistent alignment of the skeleton data across different videos, a rotation matrix was applied to standardize the orientation of the skeletons. Given the varying camera angles in the sourced videos, it was crucial to rotate the skeletons such that all poses faced the same directions for accurate analysis. A rotation matrix was used to rotate the 3D pose data around Y-axis. The angle of rotation was manually adjusted for each frame to align the skeleton to a forward-facing position. This step was performed to eliminate any discrepancies caused by varying body orientations or camera perspectives, ensuring that comparisons between lifts were based on consistent reference frames.

The rotation matrix for the Y-axis was calculated as follows:

$$\mathbf{R_y}(\theta) = \begin{pmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{pmatrix} \tag{9}$$

where θ is the angle of rotation in degrees. This transformation was applied to each frame's skeleton data, which was then visually validated by plotting the rotated and original poses side by side. The rotation process helped in aligning the skeletons in a uniform manner, facilitating more accurate movement comparisons across different athletes and lifts.This lead to rotated data vectors for each video $i \in \{1, .., N\}$ and frame $f \in \{1, .., F_i\}$:

$$\begin{pmatrix} x^*_{new} \\ y^*_{new} \\ z^*_{new} \end{pmatrix} = \mathbf{R_y}(\theta) \begin{pmatrix} x_{old} \\ y_{old} \\ z_{old} \end{pmatrix} \tag{10}$$

$$\vec{X}_{rot,f,i} = (x^*_{1,f,i}, y^*_{1,f,i}, z^*_{1,f,i}, .., x^*_{17,f,i}, y^*_{17,f,i}, z^*_{17,f,i})^T \tag{11}$$

2.2.4 Trajectory smoothing

To mitigate noise and enhance the clarity of motion data, Bezier curve smoothing was applied in two stages. First, the spine line - comprising the pelvis, spine and neck joints - was smoothed to minimize extraneous depth-related fluctuations. This was achieved by treating the spine coordinates as control nodes for a Bezier curve, which generated a smooth trajectory while preserving the natural curvature of the spine. The smoothing operation was conducted frame by frame using a function that iteratively updated the spine joints with the smoothed coordinates.

Subsequently, the entire skeleton was smoothed across the temporal dimension. For each joint, its trajectory over all frames was treated as a series of control points and a Bezier curve was fitted to capture the underlying motion pattern. This temporal smoothing ensured that joint movements remained continuous and fluid across consecutive frames, reducing abrupt positional changes caused by noise. The combined approach enhanced the dataset's quality, providing a more accurate representation of the lifters' movements for subsequent analysis. Figure 5. illustrates frame skeleton data which went through the normalization process.

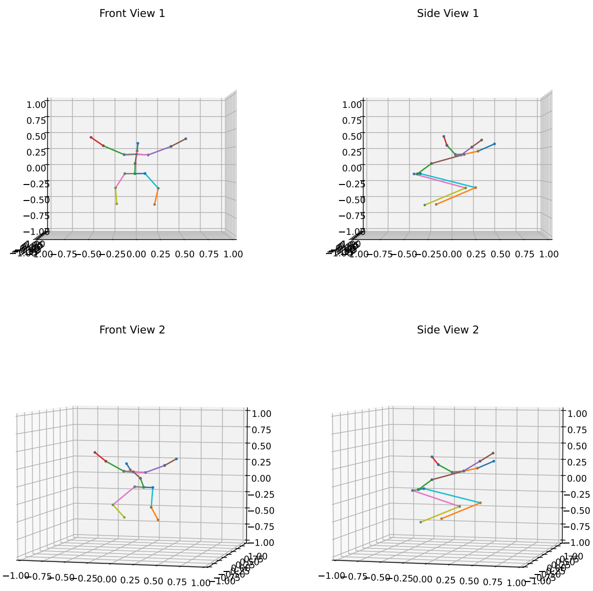

Figure 5. *Normalized skeleton data from two sides*

2.3 Golden Standard Creation

To establish the golden standard for an optimal snatch lift, the process began with visually comparing the trajectories of key joints across multiple lifts. For this, the trajectories of individual joints were analyzed in 3D space. The x, y, and z coordinates of each joint, such as the right wrist, were separated and plotted on three distinct graphs as showed in Figure 6. These coordinate-specific trajectories from multiple lifts were superimposed on the respective graphs, enabling a visual comparison of movement patterns across lifts. From this comparison, a subset of lifts that exhibited similar movement patterns was selected.

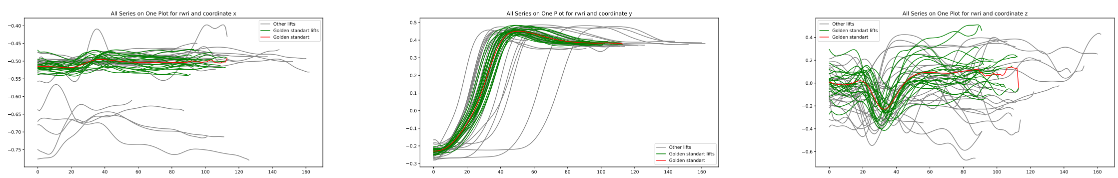

Figure 6. *Stacked graph for right wrist joint for x, y, z coordinate*

Subsequently, the selected lifts were used to calculate an average trajectory, providing a representative benchmark for the golden standard. Average trajectory was determined by using the standard average formula:

$$\bar{\mathbf{p}}_i(f) = \frac{1}{N} \sum_{j=1}^{N} \begin{bmatrix} x_{i,j,f} \\ y_{i,j,f} \\ z_{i,j,f} \end{bmatrix} \tag{12}$$

Where:

- $\bar{\mathbf{p}}_i(f) = \begin{bmatrix} \bar{x}_i(f) \\ \bar{y}_i(f) \\ \bar{z}_i(f) \end{bmatrix}$ is the average position vector of the i-th joint at frame k,
- N is the total number of high-quality lifts considered, and
- f is the index of the frame for which the average position is calculated.

To account for variations in the number of frames across different samples, shorter samples were padded with NaNs (Not a Number) to align them with longest trajectory. The average trajectory was then computed by ignoring the NaN values, ensuring that the resulting standard captures the central tendency of the selected lifts without being skewed by variable frame lengths. This golden standard serves as a benchmark for analyzing and comparing other lifts in the dataset.

2.4 Lift Comparison

For the lift comparison, Dynamic Time Warping (DTW) was employed to assess the similarity between individual snatch lifts and the golden standard. DTW is a powerful technique for measuring similarity between time series, even when they are out of phase or have different lengths. The comparison focused on the 17 key joints of the body, with each joint having 3 coordinates (x, y, z), resulting in 51 observations for each lift.

Each sample's trajectory was compared to the golden standard by calculating normalized distance between their corresponding joint coordinates (x, y, z) using the DTW method. The normalized distance is calculated for each coordinate of the joints using the formula:

$$\text{Normalized Distance} = \frac{\text{DTW Distance}}{|W|} \tag{13}$$

This formula ensures that the distance calculation is independent of the scale of the coordinates, allowing for consistent comparison across lifts, regardless of the actual joint position.

The DTW method aligns the sequences by adjusting for any temporal shifts or variations in the speed of the movement, enabling a more accurate comparison even when the samples differ in length or timing. Once the distance for all 51 columns (representing the 17 joints and 3 coordinates per joint) were calculated, the results were stored in a data frame with 53 columns (51 columns for the joint coordinates, 1 for the sample ID and 1 for metadata indicating if lift was technical correct or data contains a lot of noise). Each column represents the normalized distance for a specific joint and coordinate in a given sample.

A smaller distance indicates that the sample's trajectory is more similar to the golden standard, while a large distance suggests greater deviation from the optimal technique. This enables the identification of the lifts that are mostly outliers compared with the golden standard.

2.5 Outliers detection

In this study, we constructing the one-class classifier from good lifts as maon class, and all outlying observations are named as not correct lifts also known as outliers. Outliers were defined as lifts with deviations from the golden standard that exceeded a pre-defined threshold, representing lifts with potentially poor technique. The identification process focused on detecting the largest 5% of normalized distances, under the hypothesis that these correspond to lifts with significant deviations in joint trajectories.

The dataset consisted of 2,346 normalized DTW distances (46 samples x 51 joint-coordinate features). To identify outliers, the distances were aggregated and a threshold was set at the 5th percentile. This approach assumes that the largest distances, representing the top 5% of deviations, correspond to lifts with poorest technique.

By examining the distribution of distances across all features and samples, a cutoff value was established for the 95th percentile. Lifts with one or more distances exceeding this threshold were flagged as potential outliers. These outliers were not excluded but were instead analyzed further to provide insights into common technical issues and deviations from the golden standard.

This method ensures that the outliers reflect lifts with potential technical deficiencies, offering a data-driven basis for analyzing and improving lift quality.

3 Experiments

The experimental framework was implemented entirely using Google Colab, a cloud-based Python environment that facilitated seamless integration of data processing, analysis and visualization tools. To ensure reproducibility all the code used in this research are made available on GitHub[1].

The workflow was divided into four primary stages, each handled by separate notebook:

- *Stage 1:* Data sourcing
- *Stage 2:* Skeleton data extraction from video footage.
- *Stage 3:* Normalization of skeleton data and creation of the golden standard.
- *Stage 4 :* Application of Dynamic Time Warping (DTW) for lift comparison and outlier detection.

This modular structure ensured efficient development and testing, allowing for easy troubleshooting and scalability. Each notebook focused on specific aspects of the analysis, streamlining the overall experiment and maintaining a clear separation on tasks.

3.1 Data Source and collection

A dataset of 46 snatch lift videos was constructed to provide a representation of athlete performance. Videos were sourced from two YouTube channels: Jėgos Sportas Tiesiogiai and Weightlifting House. The selection aimed to balance variety in skill levels, encompassing both national and international competitors.

Implementation of Video Collection The video preparation process was carried out with Shortcut software. Each video was trimmed to isolate the snatch lift execution phase. This step was done manually using timestamps annotated during video review.

Dataset Properties Videos included lifts recorded under varying conditions, such as different camera angles, lighting and resolutions. This variety added robustness to the analysis by introducing natural variability. All processed videos were stored on Google Drive, ensuring accessibility and centralized storage. To integrate these files into the Google Colab environment, public links were generated for each video.

3.2 Algorithm of the workflow

The proposed algorithm is designed to evaluate the quality of snatch lift using skeleton data extracted from video footage. Each step of the process is essential for ensuring accurate and meaningful analysis. Below, Algorithm 1 is outlined along with justification for each step:

[1] `https://github.com/hermanas0vait/master_thesis`

Algorithm 1

Input: *Collected data:* Raw video footage, serve as the primary data source for analysis.
Step 1: *Skeleton Data Extraction.* Use a pre-trained pose estimation model to extract skeleton key-points, converting video frames into structured motion data.
Step 2: *Normalization. Scaling and Centering.* Adjust the skeletons to a uniform scale and center them to eliminate size and positional variance.
Step 3: *Normalization. Rotation.* Use rotation matrix to align the skeleton data consistently.
Step 4: *Normalization. Smoothing.* Apply Bezier curve smoothing to minimize noise in the extracted data.
Step 5: *Golden Standard creation.* Construct a template from an ideal lift to serve as the benchmark for evaluation.
Step 6: *Comparison via DTW.* Utilize Dynamic Time Warping to align and compare each lift to the golden standard, ensuring flexibility in timing differences.
Step 7: *Outlier Detection.* Identify lifts that deviate significantly from the standard, highlighting errors in execution.
Output: *Technical incorrect lifts.* List of lifts with wrong technical execution.

By breaking down the algorithm into these well-defined steps, the process ensures robustness, accuracy and interpretability for analyzing snatch lifts.

3.3 Skeleton data extraction

The process of extracting skeleton data from the collected video footage was carried out in three main stages using Google Colab:

1. *Data Loading:*

   Video files were initially stored on Google Drive and made publicly accessible via shared link. Using the wget command, these videos were downloaded directly into the Colab environment for prepocessing. This approach streamlined data handling and ensured all files were readily available for further stages.

2. *Model Loading:*

   A pre-trained pose estimation model 'metrabs_xl', based on a geometry-aware autoencoder and hosted on TensorFlow Hub, was employed for skeleton data extraction. The model was specifically chosen for its ability to accurately estimate 3D poses in various conditions, including competitive sports scenarios. Once the model was loaded into the Colab environment, it was prepared to process video frames and output skeleton keypoints in the 'mpi_inf_3dhp_17' format.

3. *Frame-by-Frame Processing:*

   Each video was processed frame by frame, with the model extracting the 3D coordinates of 17 key joints (e.g., shoulders, elbows, knees) from each frame. The outputs were saved as JSON files, with one file generated per video, containing the sequential motion data for the respective lift.

3.4 Data normalization

The data normalization process, conducted in Google Colab, was a critical step to ensure consistency and accuracy across all skeleton data. This phase focused on resolving inconsistencies introduced during skeleton extraction and preparing the data for analysis.

Key steps in normalization:

1. *Handling Missing Data:*

   During skeleton extraction, some frames contained missing joint data, particularly when joints were occluded or undetected. A linear interpolation approach was applied to fill gaps, using values from adjacent frames to estimate missing positions. This ensured smooth and continuous motion data across all frames.

2. *Scale and Center Adjustment:*

   Skeletons extracted from videos varied in scale and positioning due to differences in athlete size and camera angles. To standardize this:

   - The mean of each joint's 3D coordinates was subtracted to center the data around the body.
   - Coordinates were then scaled to a consistent range of [-1, 1], ensuring comparability across videos.

3. *Rotation for Alignment:*

   Since the videos were recorded from varying camera perspectives, it was necessary to align the skeletons to a common forward-facing orientation. Using a rotation matrix applied to the Y-axis, the skeletons were reoriented. The angle for rotation was determined manually for each video to archive consistent alignment.

4. *Visualization with GIFs:*

   GIFs were generated for each lift using scaled, centered and rotated skeleton data. These visualizations provided an intuitive way to observe the joint trajectories frame by frame. It was noted that certain joints exhibited 'jumping' behavior, where their positions fluctuated erratically between frames, likely due to noise in the pose estimation process. This observation reinforced the need for a robust smoothing process.

5. *Bezier Curve Smoothing:*

   To reduce noise in the skeleton data, a two-step Bezier curve smoothing approach was applied:

   - First, the spine line (pelvis, spine and neck joints) was smoothed to minimize fluctuations.

- Next, temporal smoothing was performed across all joints to ensure fluid transitions between frames.

The normalization pipeline successfully transformed raw skeleton data into a standardized and smooth dataset. The generation of GIFs early in the process provided critical insights into the initial quality of the data, ensuring that the normalization steps effectively addressed the identified issues.

3.5 Golden Standard Creation

The creation of a golden standard was a pivotal step in the experiment, providing a benchmark for comparing individual snatch lifts. This process focused on extracting the 'optimal' movement patterns from a subset of the data and generating a reference trajectory for further analysis.

Key steps in golden standard creation:

1. *Trajectory Comparison and Filtering:* The trajectories of the lifts were compared across the key joints. The x, y and z coordinates of each joint for every frame were plotted to visualize the movements. A set of lifts that exhibited similar patterns in joint trajectories was chosen to construct the golden standard. By aggregating the most consistent lifts, the resulting golden standard would represent an 'ideal' snatch lift.

2. *Handing Variable Frame Lengths:* Since the videos varied in length (due to differences in the duration of the lifts), it was necessary to handle the mismatch in frame numbers. Shorter lifts were padded with NaN (Not a Number) values to align their length with the longest lift in the selected subset. This ensured that all lifts used in the golden standard had the same number of frames, allowing for fair averaging. During the averageing process, the NaN values were excluded, so they did not affect the final trajectory.

3. *Calculation of the Average Trajectory:* The average trajectory was computed by combining the x, y and z coordinates of the selected lifts. The joint positions across all selected lifts were averaged frame-by-frame, and this resulted in a reference trajectoy for each key joint. This process was repeated for all 17 key joints to produce a complete golden standard representing the optimal snatch lift technique.

It can be observed, that identifying the 'ideal' lifts for the golden standard was a subjective process, relying heavily on visual inspection. Although this provided an effective way to select representative lifts, there was an element of human judgment involved. This could introduce some bias, though the approach was based on clear visual cues, such as joint trajectory consistency and fluid motion.

3.6 Lift Comparison

Once the golden standard was established, the next step in the experiment was to compare the individual snatch lifts to the golden standard using Dynamic Time Warping (DTW). This comparison allowed us to assess how closely each lift aligned with the ideal movement pattern, enabling the identification of deviations and potential areas for improvement.

The key steps in lift comparison:

1. *DTW Implementation:* Dynamic Time Warping (DTW) from the package dtw-python[6] was employed as the primary technique for comparing each individual lift to the golden standard. In this study, the joint trajectories of each lift were treated as time series, where each joint's position (x, y, z) over time was analyzed.

   Figure 7. illustrates good and bad samples comparison to the golden standard in a sense of the trajectory. Additionally alignment path between points of two sequances are visulaized.

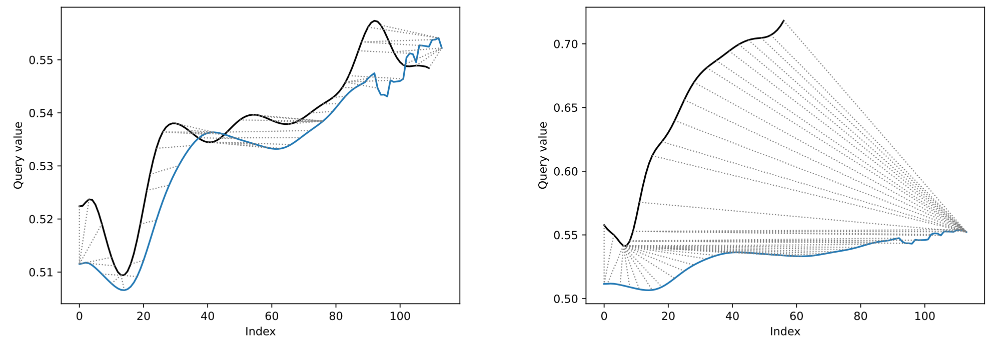

Figure 7. *Two-way comparison of the left wrist x-coordinate for good (left) and bad (right) lifts, highlighting differences in movement patterns.*

Figure 8. illustrates good and bad samples comparison to the golden standard. It shows the differences in the cost matrices and warping paths.

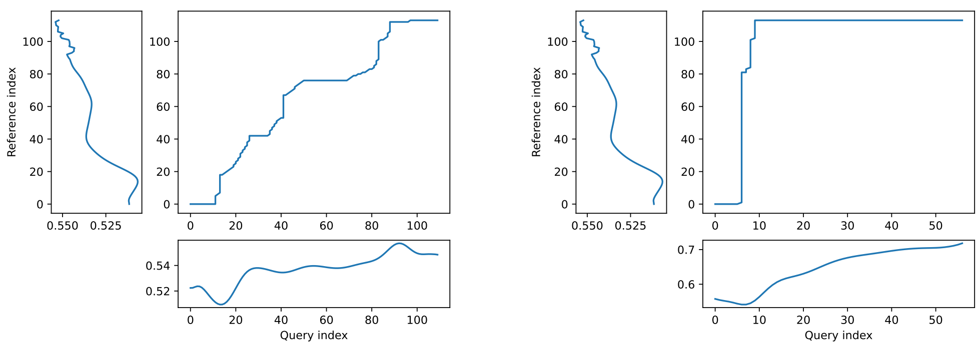

Figure 8. *Three-way comparison of the left wrist x-coordinate for good (left) and bad (right) lifts, highlighting differences in movement patterns.*

2. *Distance Calculation:* For each lift, the DTW algorithm compared the coordinates of 17 key joints (each with x, y and z coordinates, resulting 51 features per lift) to the corresponding coordinates in the golden standard. DTW calculated the optimal alignment between the sequences, allowing us to determine how similar or different each lift was compared to the golden standard. The results were then stored in a data frame, which contained the normalized distances for all 51 features, as well as metadata indicating whether the lift was technically correct/incorrect or contained a lot of noise.

Figure 9. illustrates the distribution of generated normalized distances for each coordinate. This visualization helps identify which coordinate contributes most to potential outliers.

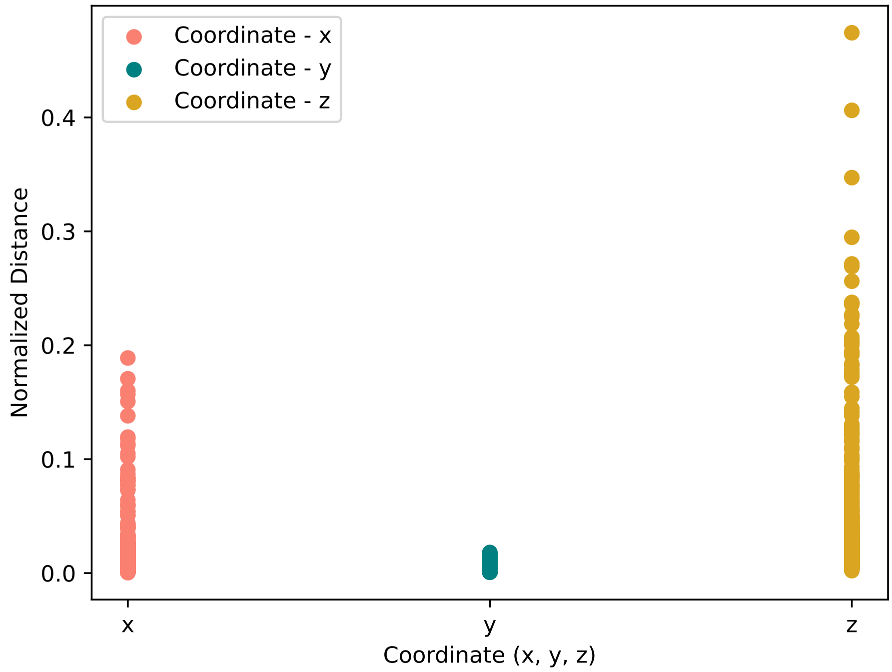

Figure 9. *Distribution of Normalized Distances by Coordinate*

3.7 Outliers Detection

The final stage of the experiment involved identifying lifts with significant deviations from the golden standard. The deviations, represented by unusually high normalized DTW distances, were classified as outliers. The goal of this phase was to systematically detect lifts that demonstrated technical flaws or irregularities in their execution.

Key steps in outliers detection:

1. *Threshold Definition:* To establish a threshold for identifying outliers, the distribution of normalized DTW distances was analyzed across all lifts and features. Using a percentile-based approach, the threshold was set at the 95th percentile of the distance distribution. This threshold captured the largest 5% of normalized distances, corresponding to the lifts

with the most significant deviations from the golden standard. This approach ensured that outliers were defined based on the data's natural variation, proving an objective basis for their detection.

2. *Outliers Identification:* For each lift, the normalized DTW distances of all 51 joint-coordinate features were compared against the threshold. Lifts with one or more distances exceeding the threshold were flagged as potential outliers. This method allowed for the identification of lifts with specific joint deviations as well as those with overall poor alignment to the golden standard.

3. *Visualization of Outliers:* To better understand the nature of the detected outliers, visualizations of normalized distances were generated as showed in Figure 10. These visualizations provided insights into the specific areas of deviation, such as irregular arm movement or improper hip alignment, offering a clearer understanding of the technical issues present in each flagged lift.

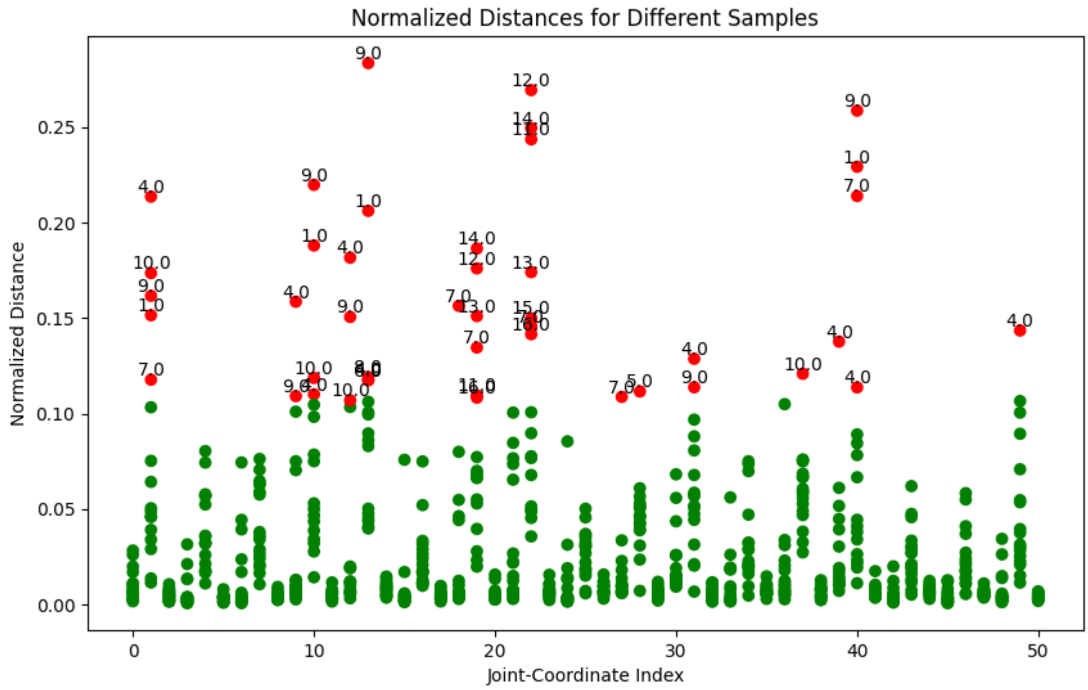

Figure 10. *Plotted potential outliers*

4. *Documentation and Analysis:* All flagged lifts were stored in a results table, which included:
   - The sample ID of the lift.
   - Outlier indicator in meta data.
   - The joints and coordinates with distances exceeding the threshold.
   - The maximum normalized DTW distance for each flagged feature.

This systematic table Table 1. allowed for a structured review of the flagged lifts, ensuring that the results could be easily interpreted.

Lift ID	Outlier indicator	Number of outliers	Maximum distance	Outlier coordinates
0	1	6	0.1381	z-6
1	1	8	0.1549	x-3; z-5
2	1	3	0.2371	z-3
3	1	5	0.1093	z-5
4	1	11	0.2042	x-5; z-6
5	1	4	0.2358	z-4
6	1	5	0.2691	z-5
7	1	6	0.1506	x-5; z-1
8	1	3	0.1826	z-3
9	1	11	0.1937	x-4;z-8
10	1	11	0.4062	x-3;z-8
17	0	3	0.1223	z-3
18	0	3	0.2562	z-3
19	0	1	0.0713	z-1
20	0	4	0.2186	z-4
21	0	11	0.4743	x-4; z-7
22	0	4	0.2072	z-4
23	0	6	0.1946	z-6
24	0	3	0.1281	z-3
25	0	3	0.3472	z-3
29	0	2	0.1017	z-2
31	0	2	0.115 4	z-2
39	0	2	0.0815	z-2
45	0	1	0.0726	z-1

Table 1. *Detected potential outliers*

3.8 Error analysis

The error analysis provides insights into the performance and reliability of the proposed one-class classification framework. The dataset consisted of 46 observations, with 16 classified as having a bad technique and 26 classified as having a good technique. This distribution highlights the inherent imbalance in the dataset, reflecting the real-world tendency for athletes to exhibit varied levels of technical proficiency.

Classification table

A confusion matrix summarizing the classification results is presented in Table 2. It outlines the true positives, false positives, true negatives, and false negatives, providing a comprehensive view of the classifier's performance.

Actual/Predicted	Good Lift	Bad lift
Good lift (29)	16	13
Bad lift (16)	6	10

Table 2. *Confusion matrix of results*

From the table

- True Positives (Good Lifts Correctly Classified): 16
- False Negatives (Good lifts Misclassified as Bad): 13
- True Negative (Bad lifts Correctly Classified): 10
- False Positives (Bad Lifts Misclassified as Good): 6

The results demonstrate that the classifier shows moderate accuracy in identifying lifts with proper technique but faces challenges in consistently distinguishing between good and bad lifts. A significant number of good lifts were misclassified as bad, and some flawed lifts were identified as good, highlighting areas for potential improvement.

Figure 11. illustrates true positive lift. As expected visualized trajectories of the right hip movement in z-coordinate aligns with golden standard.

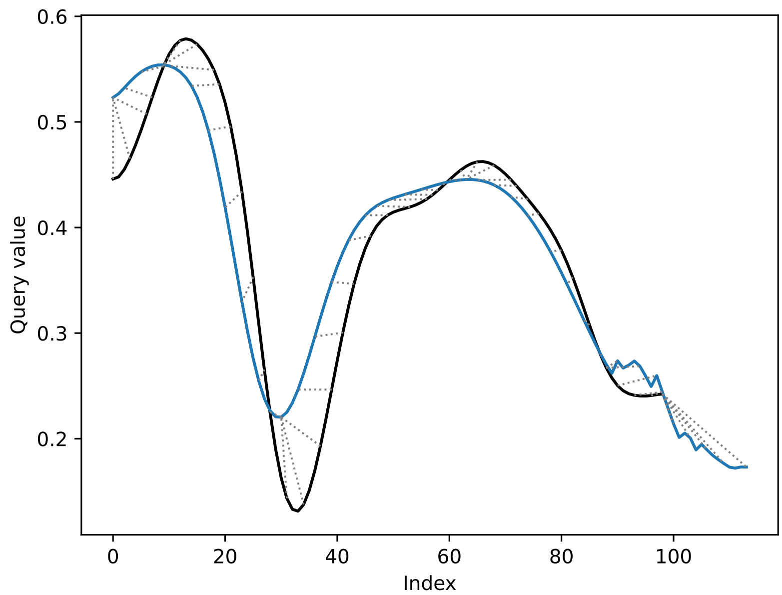

Figure 11. *Two-way graph illustrating the right hip's z-coordinate trajectory for a true positive lift.*

Figure 12. illustrates false negative lift. From the figure we can see that this could be caused by the DTW algorithm, which didn't noticed the correct alignment path between points.

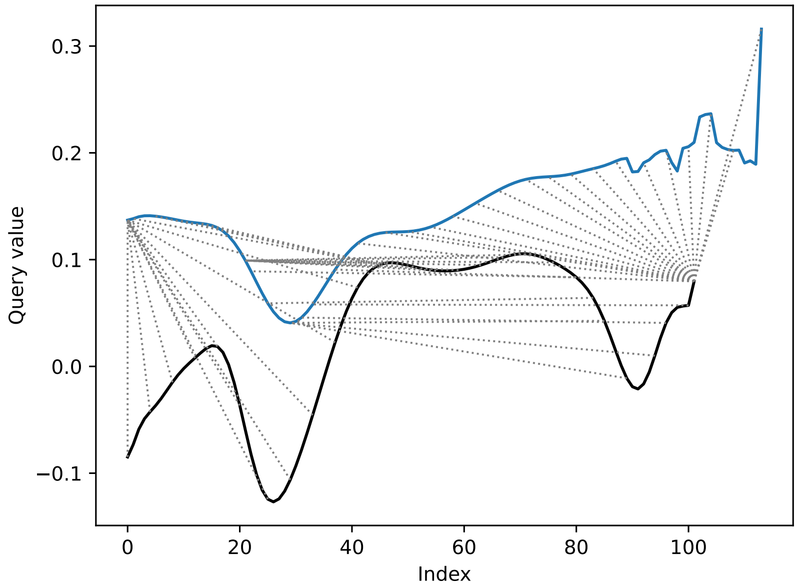

Figure 12. *Two-way graph illustrating the left ankle's z-coordinate trajectory for a false negative lift.*

Figure 13. illustrates true negative lift. It correctly indicated the error in the left wrist movement in the lift.

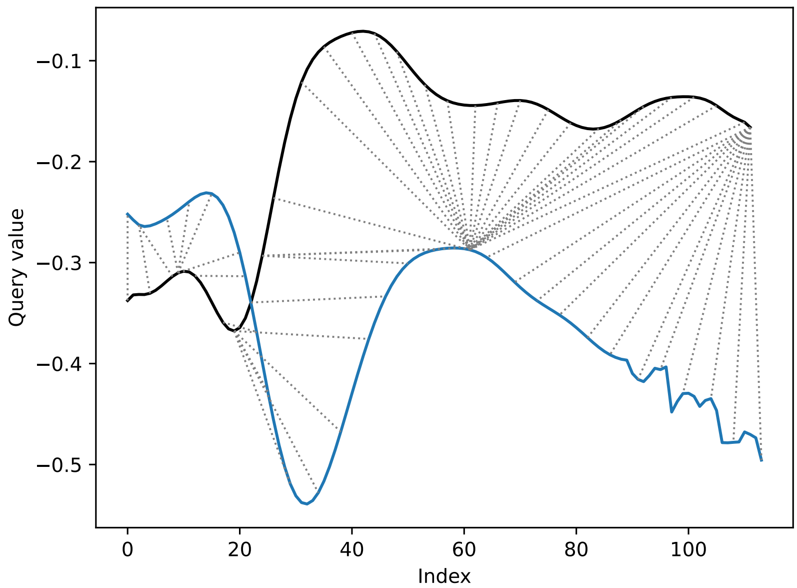

Figure 13. *Two-way graph illustrating the left wrist's z-coordinate trajectory for a true negative lift.*

Figure 14. illustrates false positive lift. It was indicated this way due to the fact, that the error, during a lift, was not accurately observed in the z-coordinate.

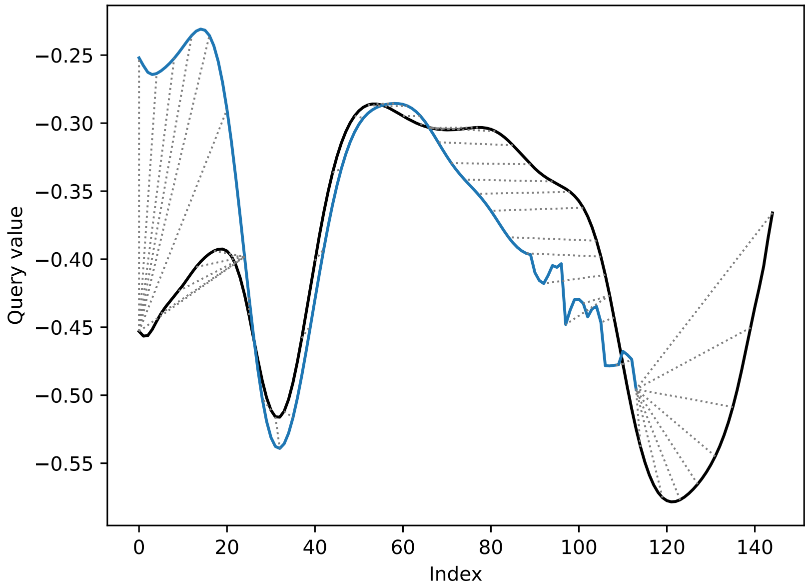

Figure 14. *Two-way graph illustrating the left wrist's z-coordinate trajectory for a false positive lift.*

Future directions and improvements:

- *Expanded Dataset:* Increasing the dataset size, particularly for bad lifts, would improve the classifier's robustness and reduce potential bias from data imbalance.
- *Multiview Analysis:* Incorporating additional camera angles would enhance the accuracy of skeleton data extraction, mitigating the impact of noisy depth coordinates and occlusions.
- *Advanced Techniques:* Experimenting with advanced methods for golden standard creation might enhance the classification capabilities.

4 Results

The outcome of this research, aligned with the defined objectives, are as follows:

1. Skeleton data was extracted using the 3D pose estimation model presented in *Learning 3D Humna Pose Estimation From Dozens of Datasets Using a Geometry-Aware Autoencoder To Bridge Between Skeleton Formats*. This model successfully extracted skeleton data from video recordings of snatch lifts, achieving accurate joint detection even in dynamic movements.

2. A combination of scaling, rotation matrix adjustments, and Bezier curve smoothing was applied to normalize the extracted skeleton data. This process effectively removed inconsistencies caused by frame-by-frame variations, resulting smooth and standardized motion trajectories across lifts.

3. The golden standard lift was created by averaging the joint coordinates across high-quality lifts. The golden standard represents a clear benchmark for optimal technique.

4. The DTW algorithm successfully compared individual snatch lifts to the golden standard. The normalized DTW distance values provided a quantitative metric for assessing lift quality.

   - High-quality lifts showed minimal deviations, with normalized distances consistently below a set threshold.
   - Poorer lifts exhibited larger distances, identifying deviations in technique.

5. Using DTW distance thresholds, lifts with significant deviations from the golden standard were flagged as outliers. When tested on 16 unseen videos, the method achieved and accuracy of 62%.

5 Conclusions

1. This study successfully proposed and developed a one-class correct snatch lift classification method for video records.

2. Normalization techniques, including scaling, centering, rotation, and smoothing, were crucial for ensuring consistent alignment of joint data across different lifts.

3. The creation of a "golden standard" was effective in establishing a reference for optimal snatch lift execution.

4. Dynamic Time Warping (DTW) proved to be reliable method for handling temporal variations in lift execution, aligning key phases of the lift despite differences in speed or duration between the query and reference sequences.

5. The proposed framework demonstrated the ability to identify outliers effectively, which can be used for detecting irregular lifts or potential errors. However, its accuracy is highly dependent on the quality and diversity of the data used to establish the golden standard, underscoring the need for further validation with more extensive and varied datasets.

6. The depth (z-axis) coordinate of the skeleton data was notably noisy, due to reliance on single-camera video recordings. This highlights the potential benefit of multi-camera setups or depth sensors to improve 3D pose accuracy and framework precision.

References

[1] B. Artacho, A. Savakis. *UniPose: Unified Human Pose Estimation in Single Images and Videos*. 2020. URL: `https://arxiv.org/abs/2001.08095`.

[2] B. Artacho, A. Savakis. *OmniPose: A Multi-Scale Framework for Multi-Person Pose Estimation*. 2021. URL: `https://arxiv.org/abs/2103.10180`.

[3] B. Artacho, A. Savakis. "UniPose+: A Unified Framework for 2D and 3D Human Pose Estimation in Images and Videos". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.12 (2022), pages 9641–9653. `https://doi.org/10.1109/TPAMI.2021.3124736`.

[4] S. Baydas, B. Karakas. "Defining a Curve as a Bézier Curve". In: *Journal of Taibah University for Science* 13.1 (2019), pages 522–528. `https://doi.org/10.1080/16583655.2019.1601913`. URL: `https://doi.org/10.1080/16583655.2019.1601913`.

[5] J. Faraway, M. Reed, J. Wang. "Modelling Three-Dimensional Trajectories by Using Bézier Curves with Application to Hand Motion". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 56 (2007), pages 571–585. `https://doi.org/10.1111/j.1467-9876.2007.00592.x`.

[6] T. Giorgino. "Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package". In: *Journal of Statistical Software* 31 (2009), pages 1–24.

[7] G. P. L. McCarthy. *Common Weightlifting Injuries*. URL: `https://www.mygcphysio.com.au/education/common-olympic-weightlifting-injuries` (viewed 2025-01-03).

[8] C. Moss. *A technical breakdown of the snatch*. URL: `https://www.canterburystrength.com/post/snatchtechnique` (viewed 2025-01-03).

[9] I. Sárándi, A. Hermans, B. Leibe. "Learning 3D Human Pose Estimation from Dozens of Datasets using a Geometry-Aware Autoencoder to Bridge Between Skeleton Formats". In: *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 2023.

[10] A. Zingoni, M. Diani, G. Corsini. "Tutorial: Dealing with rotation matrices and translation vectors in image-based applications: A tutorial". In: *IEEE Aerospace and Electronic Systems Magazine* 34 (2019), pages 38–53. `https://doi.org/10.1109/MAES.2018.170099`.

Appendix 1. List of key joints

Joint label	Joint name
htop	Head top (the highest point on the head)
neck	Neck (the connection point between the head and torso)
rsho	Right shoulder
relb	Right elbow
rwri	Right wrist
lsho	Left shoulder
lelb	Left elbow
lwri	Left wrist
rhip	Right hip
rkne	Right knee
rank	Right ankle
lhip	Left hip
lkne	Left knee
lank	Left ankle
pelv	Pelvis (the center point of the hips)
spin	Spine (the center point of the torso)
head	Head (the center point of the head)

Table 3. *17 joints which are used in the mpi_inf_3dhp_17 skeleton format*