Application of Spatio-Temporal Graph Convolutional Networks in Strength Sports: Predicting the One-Repetition Maximum (1-RM)

Mateusz M. Kunik[1,*], Marek Kowal[1] and Krzysztof Patan[1]

[1]Institute of Control and Computation Engineering, University of Zielona Góra, ul. Szafrana 2, 65-516 Zielona Góra, Poland

Abstract

Accurate One-Repetition Maximum (1-RM) assessment is crucial in strength sports for optimizing training loads, monitoring progress, and minimizing injury risk. Traditional assessment methods, whether through direct testing or mathematical estimation, are often time-consuming, invasive, or prone to significant inaccuracies. This study proposes a novel, non-invasive approach to 1-RM prediction using only video recordings of exercise execution. By leveraging BlazePose for pose estimation and Spatio-Temporal Graph Convolutional Networks (ST-GCNs) for modeling joint dynamics, we extract a movement representation termed Performance, a combination of component and latent features indicative of physical exertion. We accurately predict each squat attempt's relative load intensity (%1-RM) based on this representation. Our method introduces a new paradigm in strength evaluation, integrating biomechanics and deep learning to enable scalable, contactless feedback in real-world training settings. To support future research, we also provide a new dataset of weighted back squats annotated with biomechanical data and metadata. To our knowledge, this is the first application of ST-GCNs to predict 1-RM in strength sports, offering a safer and more personalized alternative to conventional testing methods.

Keywords

One-Repetition Maximum, Relative Load Intensity, Spatio-Temporal Graph Convolutional Networks, Pose Estimation, Strength Sports, Computer Vision, Deep Learning

1. Introduction

In recent years, strength sports have increasingly embraced data-driven approaches to monitor, analyze, and enhance athlete performance. One of the most critical metrics in this domain is the One-Repetition Maximum, the maximal load an athlete can lift for a single repetition of a specific exercise. Accurate determination of the 1-RM plays a pivotal role in prescribing training intensities, tracking progress, and mitigating injury risk. Direct methods carry a heightened risk of injury and fatigue, while indirect approaches usually lack accuracy due to their reliance on simplified, general-purpose models.

Concurrently, the fields of computer vision and deep learning have made substantial advances in modeling human movement. In particular, graph-based neural architectures such as Spatio-Temporal Graph Convolutional Networks have demonstrated exceptional capability in capturing complex motion patterns by treating the human body as a dynamic graph of interconnected joints. These models have been successfully applied to tasks such as action recognition and rehabilitation assessment, but their application in strength training remains largely unexplored.

This study bridges that gap by introducing a novel, video-based method for predicting relative load intensity during the back squat. Our approach defines Performance as a multifaceted representation of exercise execution, combining kinematic data extracted via BlazePose and structured analysis through ST-GCN. By modeling movement quality over time, we aim to infer how close a submaximal effort is to an individual's actual 1-RM, without the need for maximal lifting attempts.

In this work, we propose the task of 1-RM prediction based purely on visual input, using spatio-temporal graph-based modeling of human motion. To support this task, we introduce a new, annotated

BEHAIV-2025: AI for understanding human behavior in professional settings, 25 October 2025, Bologna, Italy.

*Corresponding author.

M.Kunik@issi.uz.zgora.pl (M. M. Kunik); M.Kowal@issi.uz.zgora.pl (M. Kowal); K.Patan@issi.uz.zgora.pl (K. Patan)

0009-0009-1050-816X (M. M. Kunik); 0000-0002-9997-9612 (M. Kowal); 0000-0002-6989-9400 (K. Patan)

dataset of weighted back squats that combines visual pose data with contextual training information, which we will make publicly available to encourage future research. Finally, we demonstrate that our approach not only offers a safer alternative to direct testing, but also achieves superior accuracy compared to traditional estimation methods.

2. Background

2.1. One-Repetition Maximum Overview

The One-Repetition Maximum is a parameter primarily used in strength sports that defines the maximum load a person can lift in a single repetition of a given exercise [1, 2]. It serves as a key indicator for assessing muscular strength and monitoring training progress. Furthermore, accurate measurement allows for adjusting training intensity to individual capabilities and training goals. This approach ensures optimal muscle stimulation while minimizing the risk of injury [3, 4, 5].

The 1-RM is used primarily in sports such as powerlifting, Olympic weightlifting, and strongman competitions. Additionally, it is also utilized in the physical preparation of athletes in disciplines such as athletics, team sports, and combat sports. In these disciplines, high levels of muscular strength may contribute to improved speed, explosiveness, and overall physical performance [6, 7].

The 1-RM can be assessed in two ways: directly - by performing a maximum load test, or indirectly - by estimating it using submaximal loads, i.e., sufficiently heavy but below the maximal capacity.

2.2. Direct Method

In the environment of strength athletes and enthusiasts, the 1-RM is most commonly determined directly by performing a maximum load test. Due to the nature of this measurement, the method is also referred to as a trial-and-error approach.

The procedure for a 1-RM test follows a standardized protocol. Initially, the participant performs a general warm-up tailored to their individual needs and musculoskeletal capabilities. This is followed by a specific warm-up in the target strength exercise (e.g., the back squat). In subsequent attempts, the participant gradually increases the load, performing increasingly heavier sets with a decreasing number of repetitions. Initially, sets may consist of 3–5 repetitions with moderate weights, while in the final sets only a single repetition is performed. Several minutes of rest are taken between attempts to allow for full muscle recovery. The test is concluded when the participant is unable to complete a single repetition correctly. The highest load lifted with proper technique is recorded as the 1-RM result.

The traditional 1-RM test is considered the most accurate method for assessing an individual's maximal strength and is relatively easy to implement under controlled conditions. However, it also comes with notable drawbacks: it increases the risk of injury places significant strain on the nervous system, and can negatively impact health in underprepared individuals. Additionally, the procedure is time-consuming, requires careful planning and warm-up, and may be psychologically demanding, especially for less experienced athletes.

2.3. Indirect Method

An alternative to the 1-RM test is the use of indirect methods, which aim to assess an athlete's muscular strength using submaximal loads. Estimation of the 1-RM parameter is based on a low-complexity mathematical model that describes the relationship between load, number of repetitions, and maximal muscular strength.

Initially, the testing procedure may resemble the direct 1-RM test — the athlete performs a warm-up and gradually increases the load while simultaneously reducing the number of repetitions. This time, however, the athlete stops at a submaximal load. With an appropriately selected weight, the athlete performs the maximum number of repetitions possible while maintaining proper exercise technique. Based on the results obtained, the 1-RM parameter is estimated using established mathematical models.

Table 1
Formulas from selected studies commonly used to estimate the 1-RM based on submaximal load (w) and number of repetitions (r) performed with that load.

Author	Formula
Epley [8]	$w(1 + \frac{r}{30})$
Brzycki [9]	$w\frac{36}{37-r}$
Lombardi [10]	$wr^{0.1}$
Naclerio et al. [11]	$w(0.951e^{-0.021r})^{-1}$
Mayhew et al. [12]	$w(0.522 + 0.419e^{-0.055r})^{-1}$
O'Conner et al. [13]	$w(1 + 0.025r)$

The indirect method estimates 1-RM using submaximal loads, offering a safer and more accessible alternative to direct testing by reducing injury risk, minimizing strain on the nervous system, and requiring less time and fewer resources. It can also ease psychological stress, as the loads used resemble regular training intensities. However, this method is generally less accurate, relying on simplified predictive models rather than actual maximal performance. Its effectiveness often depends on the athlete's training background and the specific formula used, which can limit its generalizability and introduce variability in results.

While 1-RM remains a fundamental measure of strength, both direct and indirect assessment methods have notable trade-offs in terms of accuracy, safety, and practicality. Therefore, the following section focuses on data-driven approaches and deep learning methods, which motivated us to further investigate this research problem and conduct our own analyses.

3. Related Works

3.1. Pose Estimation for Strength Sports

Markerless human pose estimation has played a key role in recent efforts to analyze athletic performance, particularly in strength training. Deep learning models such as OpenPose and BlazePose have enabled reliable extraction of skeletal keypoints from video, supporting automated analysis of exercises like the squat [14, 15, 16, 17, 18].

3.2. Graph Convolutional Networks in Human Motion Analysis

Building on this foundation, researchers have explored graph-based neural networks that move beyond raw keypoints to model the human body as a structured graph. Yan et al. [19] pioneered the Spatio-Temporal Graph Convolutional Network for skeleton-based action recognition, where each joint is treated as a graph node and spatial–temporal connections are modeled via graph convolutions. This design captures inter-joint dependencies over time, making it particularly effective for complex, coordinated motions. Compared to sequential models like LSTMs or CNNs, GCNs more effectively encode biomechanical structure and have been widely adopted for movement classification, physical therapy assessment, and rehabilitation tasks [20, 21, 22, 23]. For example, Deb et al. [23] used a variant of ST-GCN with self-attention to evaluate physical therapy exercises, outperforming prior CNN- and LSTM-based approaches on datasets like KIMORE and UI-PRMD [24, 25].

3.3. Squat Technique Analysis

A representative use case of these methods is back squat assessment, where researchers have aimed to classify technique and detect movement errors using pose data. Ogata et al. [26] proposed one of the earliest vision-based squat evaluation methods, converting 3D joint coordinates into distance matrices and analyzing them with a 1D CNN. Building on this, Youssef et al. [17] applied BlazePose and

deep learning to classify squat quality from video with high accuracy, effectively replicating coach-like evaluations based on movement cues such as knee tracking or hip depth.

3.4. Load-Velocity Relationship

Pose estimation thus provides a strong kinematic foundation for evaluating lifting mechanics. Integrating ST-GCN is a natural progression, as it enables modeling of joint coordination throughout the squat motion. Recent studies have used such models to detect errors and provide feedback in home fitness applications [27, 28]. Our work follows this trajectory, focusing specifically on back squats, and extends it by predicting 1-RM. Directly testing 1-RM can be risky or impractical, so it is often estimated from submaximal lifts. Velocity-based training relies on the established inverse relationship between lift velocity and load: as the lifted weight increases, movement speed decreases in a predictable manner [29, 30, 31]. This principle underlies many sensor and wearable-based methods that estimate 1-RM through bar speed. For example, Balsalobre-Fernández et al. [32] demonstrated that both linear models and neural networks can accurately predict bench press 1-RM from a few submaximal lifts. Recently, smartphone-based video systems have achieved comparable accuracy to hardware-based sensors when tracking barbell motion during squats [29]. Despite these advances, most indirect 1-RM prediction approaches rely on isolated features such as bar velocity or repetition count, without leveraging full-body movement data. Our method addresses this gap by using spatio-temporal pose sequences—extracted with BlazePose and modeled via ST-GCN, to estimate the relative intensity of squat attempts based on whole-body motion patterns.

4. Proposed Method

4.1. Performance

The starting point for improving the quality of 1-RM estimation lies in prediction based on Performance, understood as the execution pattern of a full set or a single repetition of a selected exercise.

We define Performance as a set of features that shape the overall quality of the attempt and provide insight into an athlete's maximal muscular strength. These features can be categorized into two groups: component and latent. Component features refer to the externally visible aspects of exercise execution, such as movement velocity, body trajectory, and the athlete's stability during the lift. These are typically extracted from video footage and represent the kinematic properties of the motion. In contrast, latent features encompass internal or non-visible aspects that influence Performance, such as training experience, athlete's skill level, or the specifics of their training program. These are generally obtained through a personal data acquisition form and are not directly inferable from visual observation.

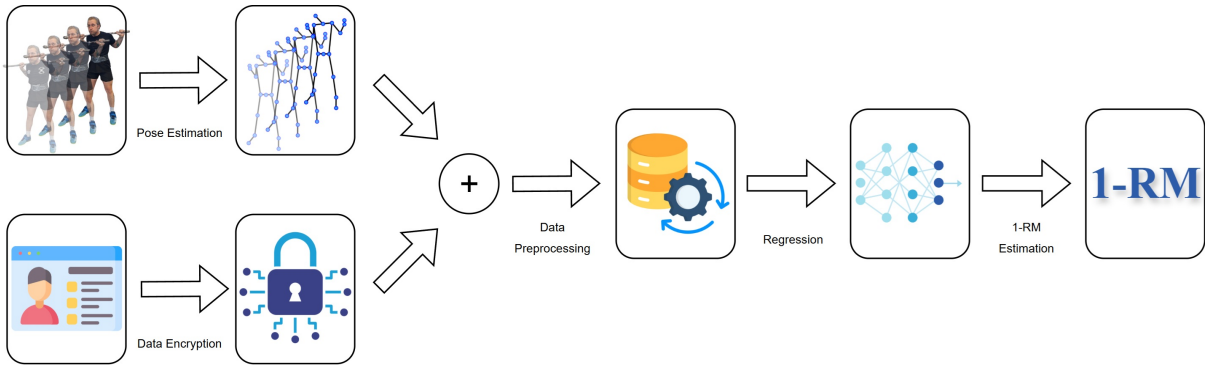

Figure 1: Visualization of the proposed 1-RM prediction pipeline based on submaximal Performance. Pose data from video and encrypted personal data are preprocessed and passed to a regression model based on spatio-temporal graph convolutions, producing an estimate of the 1-RM.

4.2. Methodology Workflow

Performance analysis requires a well-structured and carefully designed workflow, divided into several distinct stages: Below, we present our proposed sequence of steps, with the complete workflow illustrated in Figure 1.

1. **Data Collection**: Collection of video recordings of the athlete performing sets of repetitions with submaximal loads - component features. Additional personal information is gathered via a personal data acquisition form - latent features.
2. **Raw Data Processing**: The video recordings are processed using the BlazePose model [16] to estimate body posture. The resulting data represent joint coordinates changing over time during the exercise. In parallel, personal data are encrypted to ensure privacy and security.
3. **Data Concatenation**: All acquired data are concatenated into a unified structure to enable further processing.
4. **Data Preparation**: All inputs are transformed into a consistent numerical format and imputed where necessary. The data are then normalized and structured according to the input requirements of the prediction model.
5. **Model Inference**: A model based on ST-GCNs analyzes the structured data and predicts the relative load intensity, i.e., the percentage of One-Repetition Maximum lifted by the athlete in the given recording.
6. **1-RM Estimation**: Based on the predicted relative intensity and the actual weight used during the recorded attempt, the athlete's 1-RM value is estimated.

4.3. Pose Estimation

A key component of the proposed method is accurate human pose estimation from video, which serves as the basis for analyzing movement patterns during strength exercises. We employ the BlazePose model, following the findings from recent studies [16, 18, 17], which by default extracts 33 anatomical landmarks per frame. This markerless solution is optimized for real-time applications and provides a reliable skeletal representation of the athlete during the squat. The topology of the default 33 keypoints is illustrated in Fig. 2 a.

BlazePose employs a two-stage architecture consisting of a detector and a tracker, optimized for high-throughput, low-latency inference. The detector locates the full-body region of interest, while the tracker estimates landmark positions using a lightweight regression-based model. The system combines heatmap-based localization for improved spatial accuracy with direct coordinate regression to maintain speed. This hybrid approach enables robust performance even on mobile and edge devices, making BlazePose particularly well suited for biomechanical applications in real-world training environments.

In our pipeline, BlazePose processes each frame to generate a sequence of 3D coordinates, yielding a spatio-temporal tensor of shape $T \times N \times C$, where T is the number of frames, N the number of keypoints, and C the coordinate dimensions. These outputs serve as the structural foundation for graph-based modeling, enabling the next stage: relational and temporal analysis of joint movements using ST-GCN.

4.4. Spatio-Temporal Graph Convolutional Networks

To analyze full-body movement during a lift, we adopt Spatio-Temporal Graph Convolutional Networks as introduced by [19]. This model treats the human body as a dynamic graph, where joints are nodes and anatomical connections are edges. Temporal edges capture motion across time, allowing for simultaneous modeling of spatial relationships and movement dynamics.

As demonstrated in [23, 17], ST-GCN models outperform sequential or CNN-based approaches in motion quality assessment due to their ability to represent joint connectivity explicitly. Motivated by these studies and supported by initial experiments with various architectures, we selected the ST-GCN-based approach as the primary framework for our method. In our implementation, the model is tailored for regression, predicting the relative intensity of each lift based on pose sequences.

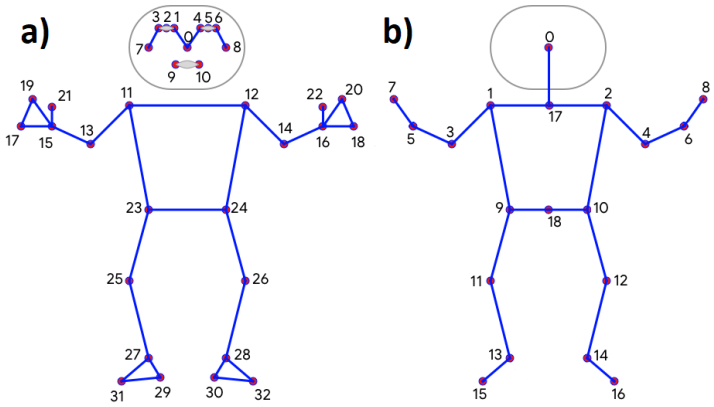

Figure 2: The figure shows the topology of keypoints: the 33 default keypoints used by the BlazePose model [16] (a); on the right, the 19 customized keypoints used in the experiment (b).

This graph-based approach ensures robust capture of movement patterns critical for estimating exertion level, offering a natural fit for analyzing complex multi-joint actions like the back squat.

5. Dataset

To effectively analyze Performance and predict the 1-RM parameter, a deep neural network must be trained on a suitably designed dataset. However, there are significant challenges in this regard. First, as we have shown, there is a lack of existing research that explores 1-RM prediction using neural networks based on video recordings. Second, the assumptions required to construct a reliable dataset for this task are difficult to meet. Consequently, no publicly available datasets fully satisfies the needs of this study. For these reasons, we were compelled to develop and annotate our own dataset of weighted back squats. We intend to release this dataset publicly to facilitate future research on video-based strength estimation.

5.1. Data Collection

Data were collected from two primary sources: video recordings and a personal data acquisition form. Video data provided observable, kinematic features of movement execution, while the questionnaire captured latent variables such as training experience and background information. The video recordings were obtained during maximum effort back squat sessions, and the forms included demographic and training-related details .

A total of 15 volunteers participated in the study, including 11 males and 4 females. The collected data span a diverse range of attributes such as age, sex, height, body weight, training experience, strength level, equipment accessibility, training program type, weekly training frequency, and participation in powerlifting competitions.

The video recording protocol adhered to standard 1-RM testing procedures, as described in Section 2. Each participant completed warm-up sets, progressively heavier squats, and single-rep attempts up to failure, defined as an inability to complete a repetition with proper technique. This generated a rich dataset of annotated squat recordings, including both submaximal and maximal efforts. After a short rest, participants also performed an AMRAP (As Many Repetitions As Possible) set at 75% of their current 1-RM, increasing sample diversity with higher-repetition submaximal examples. These sets additionally served as a dedicated test subset for evaluating traditional indirect estimation methods, with results summarized in Table2.

From a technical standpoint, each squat session was recorded using three cameras placed at fixed positions: front view, left 45° angle, and right 45° angle. All cameras were mounted on tripods at hip height relative to the participant to ensure consistency and minimize distortion. Video footage was

Table 2
Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) for models from Table 1 based on AMRAP set. Lower values indicate better estimation accuracy.

Author	RMSE	MAE	R^2
Brzycki [9]	16.8074	12.2050	0.8597
O'Conner et al. [13]	11.4115	10.4050	0.9638
Lombardi [10]	10.9727	9.7550	0.9834
Epley [8]	9.3987	7.0786	0.9491
Naclerio et al. [11]	8.4966	7.8721	0.9582
Mayhew et al. [12]	**7.0359**	**6.4364**	**0.9718**

captured at a frame rate of 30 frames per second (FPS), providing sufficient temporal resolution for detailed motion analysis.

5.2. Raw Data Processing

We began the raw data processing phase by segmenting the video recordings into short clips, each containing a single repetition of the back squat. This step resulted in a total of 1322 unique samples. Following expert review, 1255 of these repetitions were deemed technically valid and met the criteria for successful execution. Only these verified attempts were included in the modeling phase for 1-RM prediction.

The next step involved applying pose estimation using the BlazePose model, as previously discussed. Given the model's default set of 33 keypoints and the specific focus of our study we simplified the keypoint structure. Multiple landmarks representing each limb and the head were aggregated into a single point per segment, as upper-limb and head movement have limited relevance to squat Performance. However, we identified a key limitation of the BlazePose model: the lack of explicit estimation for the center of the hips and the torso. To address this, we introduced two custom keypoints representing the approximate center of the pelvis and the trunk, as these areas play a crucial role in assessing squat mechanics. The topology of the resulting 19-keypoint configuration is illustrated in Fig. 2 b.

As a result, we obtained a complete dataset ready for analysis and experimentation. The dataset is publicly available at the following link. It is distributed in three formats: raw video recordings, cropped and labeled video clips, and files containing pose estimations as described in Section 4. Each version also includes encrypted participant metadata collected via the personal data acquisition form.

6. Experiment

6.1. Model Architecture

The predictive architecture developed in this study was specifically designed to estimate the relative load intensity of a back squat attempt based on both submaximal and maximal loads. The model consists of three main components: the Squat Encoder, the Context Encoder, and the Regression Head, which collectively enable the analysis of squat Performance by incorporating observable kinematic features and user-specific contextual data. The overall model architecture is illustrated in Figure 3.

The Squat Encoder serves as the core module for processing the component features of Performance, as described in Section 4. These features are extracted from pose sequences obtained using the BlazePose model, which outputs 3D keypoint coordinates for each frame of the video. Each pose sequence is represented as a spatio-temporal graph, where joints are modeled as nodes and anatomical or temporal relationships are represented by edges. This graph is then passed through a ST-GCN, which captures both spatial and temporal dependencies in the motion data.

The implemented model includes ten consecutive ST-GCN blocks with progressively increasing channel sizes of 64, 128, and 256 (Figure 3). Each block applies a spatial graph convolution followed by a

temporal convolution, allowing the network to analyze movement over an extended time window. This facilitates the detection of temporal patterns such as rhythm and control throughout the squat. To define neighborhood relations within the graph, we tested multiple partition strategies, including uniform, distance-based, and spatial configurations, using either the thorax or pelvis as the skeleton center [19]. These strategies determine how node neighborhoods are grouped during convolution, enabling the model to emphasize different anatomical or directional aspects of the movement depending on the configuration. The final ST-GCN block is followed by global average pooling, which compresses the entire sequence into a fixed-length vector of size 256. This vector serves as a compact and informative representation of the component features extracted from the movement.

In parallel, the Context Encoder processes the latent features of Performance, as defined in Section 4. Such features provide valuable context that complements the visual motion data and support the Squat Encoder. The Context Encoder is implemented as a fully connected neural network composed of three linear layers with ReLU activation functions and dropout regularization. The layers have output sizes of 64, 128, and 256, respectively (Figure 3). The final output is a vector that encodes the athlete's profile and aligns dimensionally with the output of the Squat Encoder.

The output vectors from the encoders are concatenated and passed to the Regression Head, which performs the final prediction of relative intensity. The Regression Head consists of three fully connected layers with 1024, 512, and 64 neurons, respectively, each followed by ReLU activations and dropout. A final linear layer produces a single scalar value corresponding to the predicted %1-RM for the analyzed squat attempt.

6.2. Loss Function and Training Algorithm

The model was trained in a supervised regression setting. To guide the learning process, we employed the Root Mean Square Error (RMSE) as the loss function, due to its sensitivity to larger errors and its alignment with the objective of minimizing prediction deviation. In addition to the training loss, two evaluation metrics—Mean Absolute Error (MAE) and R-squared (R^2) were tracked to provide a more comprehensive assessment of model performance across experiments.

During the initial phase of model development, three optimization algorithms were evaluated: Stochastic Gradient Descent (SGD), Adam, and AdamW [33]. Preliminary experiments indicated that AdamW consistently achieved superior performance in terms of both convergence rate and validation error. This outcome is consistent with recent empirical studies demonstrating the advantages of AdamW over other optimizers, particularly due to its decoupled weight decay regularization and improved generalization capabilities in deep neural networks. In light of these findings and the observed empirical performance, AdamW was selected as the optimization algorithm for all experiments reported in this study.

6.3. Data Splitting and Cross-Validation Strategy

To reliably assess the quality of the trained models, we adopted the assumption that any split of a given dataset must take repetition structure into account. This means that samples originating from the same repetition, recorded by different cameras, must be assigned to the same subset. This ensures that no data leakage occurs between any two subsets. However, due to the limited number of participants, it was not feasible to perform a split into train and test sets based on individual subjects.

Due to the relatively small size of the dataset for training deep neural networks, we had to limit the test set to a sufficient minimum. Therefore, we adopted a 9:1 split ratio between the training and test sets. This corresponded to 1129 samples in the training set and 126 samples in the test set. During model training, we applied 4-fold cross-validation. The training set was divided into four subsets, three of which served as a temporary training set, while the fourth was used as a temporary validation set for model evaluation during training. The test set, which contained samples unseen by the model during training, was used exclusively for evaluation with the final epoch weights as well as with the weights that yielded the best model performance.

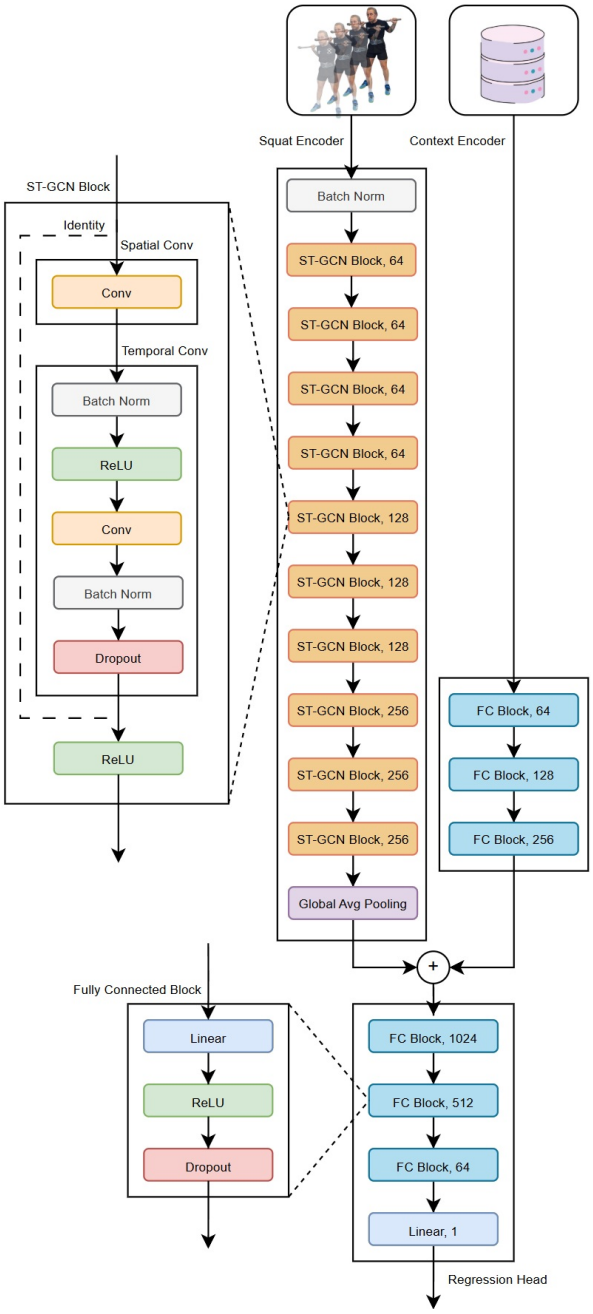

Figure 3: Architecture of the model for predicting relative load intensity. The system consists of three main modules: the Squat Encoder, the Context Encoder, and the Regression Head, which fuses both representations to generate the final prediction.

6.4. Optimization of ST-GCN Hyperparameters

To optimize the predictive accuracy of relative load intensity estimation, we conducted a series of experiments focused on selecting the most effective structural parameters for the ST-GCN model. Given the inherent architectural complexity of Spatio-Temporal Graph Convolutional Networks and the multitude of possible configurations, we focused on two principal design variables: the temporal kernel size and the maximum hop distance. Both hyperparameters were evaluated in conjunction with various partition strategies, which govern how neighborhood information is propagated between graph nodes. Each value of the temporal kernel size and max hop distance was assessed in combination with four distinct partition strategies: uniform, distance-based, spatial with the skeleton center located at the thorax, and spatial with the skeleton center located at the pelvis.

Table 3
Mean and standard deviation of RMSE scores for different temporal kernel sizes. Each result reflects the average performance over four cross-validation folds.

		Partition Strategy			
Kernel Size	Metric	Uniform	Distance	Spatial: Thorax	Spatial: Pelvis
15	Mean	8.8572	6.7342	6.5563	6.9773
	Std	0.4292	1.2971	1.472	2.2341
31	Mean	8.5068	**6.0649**	**5.8784**	**5.8596**
	Std	1.4484	**0.7138**	**0.4917**	**0.2808**
61	Mean	6.6167	7.2605	5.8847	5.9760
	Std	0.5405	0.2656	0.3993	0.9257
75	Mean	**5.8157**	6.0758	6.31533	8.8578
	Std	**0.3972**	0.4181	0.8334	0.5278

The first stage of the experiments investigated the influence of the temporal kernel size, which defines the size of the time window used in temporal convolutions. This hyperparameter controls the model's ability to capture long-range temporal dependencies within the input pose sequences. Assuming a video frame rate of 30 FPS, we tested kernel sizes of 15, 31, 61, and 75, corresponding to approximately 0.5, 1, 2, and 2.5 seconds of motion, respectively. Each kernel size was evaluated in conjunction with the four aforementioned partition strategies. For the sake of comparability, the maximum hop distance was fixed at 1 across all strategies in this stage, as the uniform strategy is only defined for a hop distance of 1. Table 3 presents the results of these experiments. The lowest RMSE overall (5.8157, SD = 0.3972) was achieved using the uniform partition strategy with a temporal kernel size of 75. However, larger temporal kernels significantly increase inference time, which may hinder real-time or resource-constrained deployment. Among the non-uniform strategies, the most consistent and favorable performance was observed with a kernel size of 31. The best configuration in this group (RMSE = 5.8596, SD = 0.2808) was obtained using the spatial strategy centered at the pelvis. Considering the trade-off between accuracy and computational efficiency, a kernel size of 31 was selected for subsequent experiments.

In the next phase of our study, we investigated the impact of the max hop distance hyperparameter, which defines the maximum number of graph edges over which information can propagate during spatial graph convolution. In practice, this hyperparameter controls how far each node can influence others during message passing within the ST-GCN layers. Each partition strategy was evaluated with max hop distances ranging from 1 to 4. Higher values were excluded from analysis, as the underlying skeleton graph used in this work contains only 19 nodes. Beyond a certain threshold, increasing the hop distance leads to a rapid saturation of graph connectivity—diminishing the benefits of localized spatial structure and increasing computational cost without meaningful performance gain. Table 4 summarizes the results. The lowest overall RMSE (4.8342, SD = 0.5930) was achieved using the distance-based partition strategy with a max hop distance of 2. This configuration outperformed all other combinations across the evaluated strategies. For the spatial partition strategy (both thorax- and pelvis-centered), the best performance was observed at a hop distance of 1, although the corresponding RMSE values (5.3674 and 5.6552, respectively) remained higher than the optimal distance-based configuration.

Based on the results of the architectural experiments, the optimal ST-GCN configuration was determined to include a temporal kernel size of 31, the distance-based partition strategy, and a maximum hop distance of 2. This setup provided the most favorable balance between prediction accuracy and computational efficiency across tested variants. Accordingly, this configuration was adopted as the default in all subsequent experiments. Furthermore, the model achieving the best performance (RMSE = 4.8342, SD = 0.5930) with these optimal settings was designated as the baseline for the remainder of the experimental study.

All architecture-related experiments were conducted under fixed training conditions. To ensure a fair comparison between configurations, all models were trained using identical hyperparameters,

Table 4
Mean and standard deviation of RMSE scores for different maximum hop distances. Each result reflects the average performance over four cross-validation folds.

		Partition Strategy			
Max Hop	Metric	Uniform	Distance	Spatial: Thorax	Spatial: Pelvis
1	Mean	**8.5068**	5.8348	**5.3674**	**5.6552**
	Std	**1.4484**	0.4785	**0.5932**	**1.1292**
2	Mean	-	**4.8342**	6.7283	6.0282
	Std	-	**0.5930**	0.3897	0.1561
3	Mean	-	5.3242	7.7605	5.9648
	Std	-	0.3812	0.4970	0.2814
4	Mean	-	6.1665	6.9031	8.4139
	Std	-	0.5363	0.6973	0.3572

which remained unchanged throughout this phase of the study. The only variables modified were the structural parameters under investigation—namely, the temporal kernel size, the partition strategy, and the maximum hop distance. Each model was trained for exactly 100 epochs, providing consistent training duration across all configurations. This setup allowed us to isolate the effect of architectural design choices on prediction performance, without introducing confounding factors from optimization dynamics.

6.5. Model Training

The proposed was trained for a maximum of 200 epochs. Given the relatively small dataset, batch sizes ranging from 16 to 64 were explored. A conventional early stopping strategy was employed to ensure training stability and prevent overfitting by halting training when validation performance ceased to improve over a specified window. Model checkpoints were monitored throughout training and saved whenever the validation loss improved. However, only the best-performing checkpoint i.e., the one achieving the lowest RMSE on the validation set—was retained for final evaluation. Learning rate scheduling was applied using strategies such as the StepLR scheduler to promote convergence. Throughout training, model performance was continuously evaluated on the validation set using the RMSE metric. After training completion, both the final epoch weights and the best checkpoint were evaluated on the test set using RMSE, MAE, and R^2.

Data augmentation was applied exclusively to the training set and limited to the input of the Squat Encoder, which processes pose data derived from BlazePose. Specifically, Gaussian noise was added to the 3D keypoint coordinates. The noise was drawn from a normal distribution with a defined mean and standard deviation. To introduce variability while preserving data diversity, this transformation was applied probabilistically, ensuring that only a portion of the samples was augmented within each training epoch.

Table 5
Performance comparison of the Mayhew model (the strongest traditional baseline on this dataset), the baseline model, and the proposed model in predicting 1-RM.

Model	Metric	RMSE	MAE	R^2
Mayhew et al. [12]	Mean	7.0359	6.4364	**0.9718**
	Std	-	-	-
Baseline Model	Mean	4.8342	3.7759	0.9361
	Std	0.5930	0.5616	0.0163
Proposed Model	Mean	**4.5412**	**3.3338**	0.9522
	Std	**0.2965**	**0.1892**	0.0063

The final model was trained using a batch size of 64, with a learning rate of 0.0001 and a weight decay of 1e-6. Dropout regularization was set to 0.1 across fully connected layers, and training was conducted for 150 epochs. A StepLR learning rate scheduler with a step size of 50 and decay factor of 0.5 was used to refine convergence. Noise-based data augmentation was applied to the pose input with a probability of 0.9, using zero-mean Gaussian noise with a standard deviation of 0.01. The ST-GCN configuration included a distance-based partition strategy, a maximum hop distance of 2, and a dilation factor of 1.

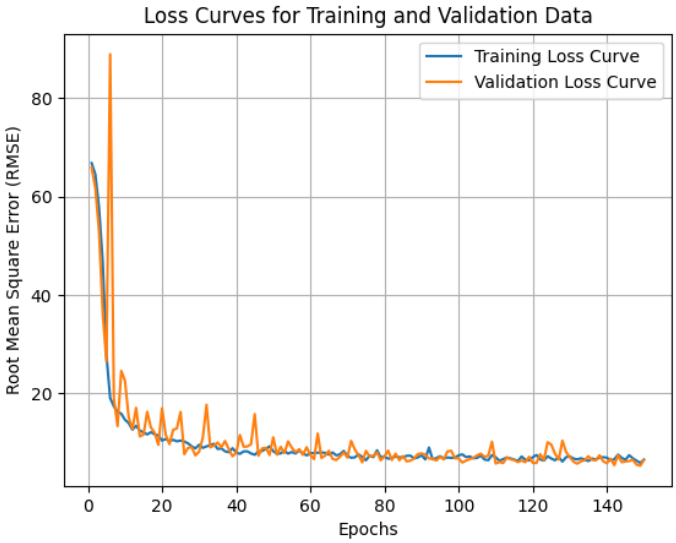

Figure 4: Training and validation loss curves over 150 epochs, measured in RMSE. The plot illustrates effective convergence of the model.

The proposed model achieved the highest predictive accuracy across all evaluated metrics, as illustrated in Table 5, outperforming both the baseline and the classical Mayhew model. It obtained the lowest RMSE (4.5412) and MAE (3.3338), with notably low variance, indicating stable performance. While the Mayhew model showed a slightly higher R^2, this can be attributed to the limitations of R^2 in nonlinear settings—it favors models that explain overall variance, whereas our model captures fine-grained, nonlinear patterns that yield more precise predictions, better reflected by RMSE and MAE.

The training curves, shown in Figure 4, confirm the model's effective convergence and generalization. Overall, the proposed model demonstrates superior performance and stability, offering a more accurate and nuanced alternative to traditional linear methods in estimating 1-RM.

6.6. Hardware and Software Configuration

The experiments were carried out on a system running Windows 11, equipped with an NVIDIA RTX 6000 Ada Generation (48 GB VRAM), an AMD Ryzen 9 7950X 16-Core Processor, and 64 GB of RAM, providing a robust environment for deep learning tasks. Python 3.11 was used as the primary programming language, with PyTorch 2.6.0+cu118 serving as the core deep learning framework.

7. Conclusion

This study presented a novel, vision-based approach for predicting One-Repetition Maximum in strength sports, specifically targeting the estimation of relative load intensity in the back squat using spatio-temporal pose data. By combining BlazePose-based pose estimation with Spatial-Temporal Graph Convolutional Networks, the method successfully inferred %1-RM from full-body movement patterns, offering a more accurate, non-invasive alternative to traditional sensor-based or formulaic methods.

The findings highlight the potential of data-driven, movement-based performance modeling, particularly in settings where safety and scalability are essential. Future research will aim to improve model generalizability through expanded datasets and enable real-time deployment in training environments using lightweight architectures. These advancements could pave the way for intelligent coaching systems that provide immediate, personalized feedback to athletes and coaches.

Declaration on Generative AI

During the preparation of this work, the authors used o4-mini for Generate literature review and Abstract drafting. Furthermore, GPT-4.1 was used for Text translation, Paraphrase and reword, Improve writing style and Grammar and spelling check. After using these tool, the authors reviewed and edited the content as needed and take full responsibility for the final version of the publication.

References

[1] R. Marchese, A. Hill, The Essential Guide to Fitness for the Fitness Instructor, Pearson Education, Frenchs Forest, N.S.W., 2005.

[2] G. G. Haff, N. T. Triplett, Essentials of strength training and conditioning 4th edition, Human kinetics, 2015.

[3] T. J. Suchomel, S. Nimphius, C. R. Bellon, W. G. Hornsby, M. H. Stone, Training for muscular strength: Methods for monitoring and adjusting training intensity, Sports Medicine 51 (2021) 2051 – 2066. URL: https://api.semanticscholar.org/CorpusID:235364256.

[4] G. R. Hunter, M. S. Treuth, Relative training intensity and increases in strength in older women, Journal of Strength and Conditioning Research 9 (1995) 188–191. URL: https://api.semanticscholar.org/CorpusID:143810641.

[5] L. M. Hickmott, P. D. Chilibeck, K. A. Shaw, S. J. Butcher, The effect of load and volume autoregulation on muscular strength and hypertrophy: A systematic review and meta-analysis, Sports medicine-open 8 (2022) 9.

[6] Y. Jeong, H. pil Jun, Y.-L. Huang, E. Chang, The squat one repetition maximum may not be the best indicator for speed-related sports performance improvement in elite male rugby athletes, Applied Sciences (2023). URL: https://api.semanticscholar.org/CorpusID:266461241.

[7] L. W. Judge, J. N. Wildeman, D. Bellar, Designing an effective preactivity warm-up routine for the 1 repetition maximum back squat, Strength and Conditioning Journal 33 (2011) 88–90. URL: https://api.semanticscholar.org/CorpusID:70488651.

[8] B. Epley, Poundage chart, Boyd Epley Workout. Lincoln, NE: Body Enterprises 86 (1985).

[9] M. Brzycki, Strength testing—predicting a one-rep max from reps-to-fatigue, Journal of physical education, recreation & dance 64 (1993) 88–90.

[10] V. P. Lombardi, Beginning weight training: the safe and effective way, (No Title) (1989).

[11] F. J. Naclerio, A. Jiménez, B. A. Alvar, M. D. Peterson, Assessing strength and power in resistance training, Journal of Human Sport and Exercise 4 (2009) 100–113.

[12] J. L. Mayhew, T. E. Ball, M. D. Arnold, J. C. Bowen, Relative muscular endurance performance as a predictor of bench press strength in college men and women, The Journal of Strength & Conditioning Research 6 (1992) 200–206.

[13] R. O'Connor, J. Simmons, P. O'Shea, Weight Training Today, West, 1989. URL: https://books.google.pl/books?id=eHqfwq6o18oC.

[14] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, Y. Sheikh, Openpose: Realtime multi-person 2d pose estimation using part affinity fields, IEEE transactions on pattern analysis and machine intelligence 43 (2019) 172–186.

[15] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, et al., Mediapipe: A framework for building perception pipelines, arXiv preprint arXiv:1906.08172 (2019).

[16] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, M. Grundmann, Blazepose: On-device real-time body pose tracking, arXiv preprint arXiv:2006.10204 (2020).
[17] F. Youssef, A. B. Zaki, W. Gomaa, Analysis of the squat exercise from visual data., in: ICINCO, 2022, pp. 79–88.
[18] J.-Y. Chen, K. Lin, C.-L. Chi, C.-J. Lin, Analysis of weight-bearing squat posture based on deep learning and sports science, 2024, pp. 273–278. doi:10.1109/ICMLC63072.2024.10935099.
[19] S. Yan, Y. Xiong, D. Lin, Spatial temporal graph convolutional networks for skeleton-based action recognition, in: Proceedings of the AAAI conference on artificial intelligence, volume 32, 2018.
[20] Y. Li, Z. He, X. Ye, Z. He, K. Han, Spatial temporal graph convolutional networks for skeleton-based dynamic hand gesture recognition, EURASIP Journal on Image and Video Processing 2019 (2019) 1–7.
[21] B. Yu, H. Yin, Z. Zhu, Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting, arXiv preprint arXiv:1709.04875 (2017).
[22] A. Mohamed, K. Qian, M. Elhoseiny, C. Claudel, Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 14424–14432.
[23] S. Deb, M. F. Islam, S. Rahman, S. Rahman, Graph convolutional networks for assessment of physical rehabilitation exercises, IEEE Transactions on Neural Systems and Rehabilitation Engineering 30 (2022) 410–419.
[24] M. Capecci, M. G. Ceravolo, F. Ferracuti, S. Iarlori, A. Monteriu, L. Romeo, F. Verdini, The kimore dataset: Kinematic assessment of movement and clinical scores for remote monitoring of physical rehabilitation, IEEE Transactions on Neural Systems and Rehabilitation Engineering 27 (2019) 1436–1448.
[25] A. Vakanski, H.-p. Jun, D. Paul, R. Baker, A data set of human body movements for physical rehabilitation exercises, Data 3 (2018) 2.
[26] R. Ogata, E. Simo-Serra, S. Iizuka, H. Ishikawa, Temporal distance matrices for squat classification, in: Proceedings of the ieee/cvf conference on computer vision and pattern recognition workshops, 2019, pp. 0–0.
[27] C. Du, S. Graham, C. Depp, T. Nguyen, Assessing physical rehabilitation exercises using graph convolutional network with self-supervised regularization, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2021, pp. 281–285.
[28] J. Bai, Z. Wang, X. Lu, X. Wen, Improved spatial–temporal graph convolutional networks for upper limb rehabilitation assessment based on precise posture measurement, Frontiers in Neuroscience 17 (2023) 1219556.
[29] B. Pueo, J. J. Lopez, J. M. Mossi, A. Colomer, J. M. Jimenez-Olmedo, Video-based system for automatic measurement of barbell velocity in back squat, Sensors 21 (2021) 925.
[30] J. Weakley, B. Mann, H. Banyard, S. McLaren, T. Scott, A. Garcia-Ramos, Velocity-based training: From theory to application, Strength & Conditioning Journal 43 (2021) 31–49.
[31] S. W. Thompson, D. Rogerson, A. Ruddock, L. Greig, H. F. Dorrell, A. Barnes, A novel approach to 1rm prediction using the load-velocity profile: a comparison of models, Sports 9 (2021) 88.
[32] C. Balsalobre-Fernández, K. Kipp, Use of machine-learning and load–velocity profiling to estimate 1-repetition maximums for two variations of the bench-press exercise, Sports 9 (2021) 39.
[33] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, arXiv preprint arXiv:1711.05101 (2017).

A. Online Resources

To support reproducibility and further research, the full source code and the dataset prepared for these experiments are provided in the following repository

- GitHub