BETWEEN- AND WITHIN-DAY REPEATABILITY OF MARKERLESS 2D MOTION ANALYSIS USING DEEP NEURAL NETWORKS

Vesa Romppanen

Master's thesis
Supervisors Neil Cronin and Janne Avela
Faculty of Sports and Health
University of Jyväskylä
2021

ABSTRACT

Romppanen, V. 2021. Between- and within-day repeatability of markerless 2D motion analysis using deep neural networks. Faculty of Sport and Health Sciences, University of Jyväskylä, Master’s thesis, 39 pp.

The purpose of this study was to evaluate kinematic analysis repeatability by deep learning approach in countermovement jump. Seventy athletes (39 women, 31 men) performed two maximal countermovement jumps in either one session or two separate sessions (jumps separated by two-weeks). The jumps were filmed from lateral and frontal point of view. Video data from 50 athletes were selected randomly to be used for training the deep learning model with DeepLabCut. A total of 10 images were used from every athlete from this training set, meaning that a total of 500 images were used to create the model for frontal view and side view (sagittal) videos. The performance of this model was then evaluated by applying it on 11 within-day measurements and 9 between-day measurements again for both frontal and sagittal videos. For frontal view videos, the marker locations were labelled for both sides of the body to shoulder (acromion), hip joint (greater trochanter), knee joint (mid-point of patella) ankle joint (mid-point between malleoli) and toes (head of shoe). The marker locations of shoulder (acromion), hip joint (greater trochanter), knee joint (lateral femoral condyle), ankle joint (lateral malleolus) and toes (head of shoe) were manually labelled for sagittal test images. For the sagittal videos, hip, knee and ankle joint angles were calculated by using atan2 function in Matlab, and for the frontal view videos, the same was done for the knee and ankle angles. To compensate for misplaced or missing markers, raw data was filtered with a median filter and subsequently with Butterworth 4th order low-pass filter. After filtering, data was further processed with Matlab by first aligning the curve data of consecutive (trial 1 and trial 2) jumps. Then data was cropped according to the movement of knee joint from sagittal plane: start of cropping was selected as the point where there was a 5-degree joint angle change from the initial standing position, and the end point was selected as the same calculated value after landing the countermovement jump. Test-retest values were calculated with intraclass correlation coefficients (ICC) for subjects in the evaluation set. The ICC model used for test-retest was single measurement two-way mixed effects with absolute agreement. High mean ICC values were observed for sagittal within-day joint angles (0.95 ± 0.04 for hip joint, 0.96 ± 0.03 for knee joint and 0.95 ± 0.05 for ankle joint). Similar values were found for mean between-day measurements (0.95 ± 0.03 for hip joint, 0.95 ± 0.07 for knee joint and 0.89 ± 0.08 for ankle joint). On the contrary, correlations of joint angle values for frontal plane varied substantially more: For within-day measurements, mean ICC values revealed poor test-retest reliability for right knee angle (ICC = 0.43 ± 0.31), and moderate test-retest reliability for left knee (ICC = 0.68 ± 0.23), right ankle (ICC = 0.62 ± 0.22) and left ankle (ICC = 0.53 ± 0.29) angles. Mean between-day ICC values demonstrated good (ICC = 0.75 ± 0.10) test-retest reliability for right knee angle, moderate test-retest reliability for left ankle angle (0.53 ± 0.17), and poor test-retest reliability for left knee (ICC = 0.49 ± 0.27) and right ankle (ICC = 0.34 ± 0.26) angles. These results imply that deep learning approach provides very repeatable measurements for sagittal joint angles in countermovement jump, but not as such for frontal plane kinematics. Hence deep learning approach provides an affordable and easy-to-access method to perform repeated measurements for 2-D motion analysis of countermovement jump and possibly other sports movements filmed from sagittal plane. Further studies on repeatability and the validation of deep learning-based systems are required to prove their accuracy and to provide reliable data for practitioners.

CONTENTS

1 INTRODUCTION

Human motion analysis has progressed gradually, and it is continuously applying more sophisticated technological tools, including markerless systems that utilize human body models, computer vision and machine learning (Colyer et al. 2018). The increasing numbers of research and the evolvement of these novel kinematic analysis tools is highlighting the importance of developing and testing technology in the field and exploring the opportunities they give. Until recent years, motion analysis has relied heavily on optoelectronic measurement systems, which are often described as the gold standard for movement analysis. The basic principle of optoelectronic system is detecting light from markers, which in turn is turned to electrical signal in the camera, whereas when using computer vision based and automated systems, markers are not required and often described as "markerless" motion analysis (van der Kruk & Reijne 2018; Mathis et al. 2018) Even though optoelectronic systems have been the gold standard for kinematic analysis for few decades, systems that work by using automated detection and recognition of poses and sport specific movement are becoming more prominent in biomechanical research and are appearing as practical applications.

Using computer vision has applications for performance analysis including player tracking, semantic analysis, and movement analysis (Cust et al. 2019). Markerless systems could provide a great contender to optoelectronic performance analysis systems especially in practical aspects due to their lower cost, invasiveness and time saving aspects, such as faster subject preparation and faster feedback to the practitioners. Additionally, markerless measurements may be performed outside of laboratory, providing quick and easily accessible kinematic analysis system to its users.

One of the most common sports tests for player performance evaluation and monitoring is countermovement jump (CMJ). The reliability of CMJ performance (jump height) has been explored quite vastly, with good implications of low variability (Carrol et al. 2019). However, there is variability present in CMJ kinematics between intra-subject trial-to-trial tests, just as it is higher in many other common sports tests and sports movements with fast power production (Raffalt et al. 2016; Wren et al. 2020). This is an important factor to take into account when considering the use of sports movements as a base to measure the accuracy and repeatability of a kinematic analysis system.

Before markerless systems can be applied for biomechanical research or practical applications, their accuracy, repeatability, validity, and reliability should be confirmed. Machine learning shows possibility as method to perform relatively cheap kinematic analysis outside laboratory environment. Deep learning-based machine learning performance has been assessed for sagittal plane kinematics with underwater running, vertical jumps, squatting, walking and running (Cronin et al. 2019; Drazan et al. 2021; Ota et al. 2020; Ota et al. 2021). There is need for evaluation on the accuracy of deep learning as a motion analysis tool in recreational sports movements. Due to its prevalence in sport testing and research, and similarity to many other sports movements, CMJ is a fitting sports test to be assessed for kinematic analysis system performance. Hence the aim of this study is to evaluate kinematic analysis repeatability by deep learning approach in common sport testing situation with countermovement jump.

2 MOTION ANALYSIS

Analysing kinematic parameters derived with motion capture systems has become widely used method for biomechanical studies and applications. Earliest versions of motion analysis can be dated back to 1878, with Muybridge's sequence of photos of a galloping horse (Chiari et al. 2005). From there on, motion capture sensors and systems have gradually transformed from relatively inaccurate and time-consuming methods to more profound, practical, and accurate systems. Motion camera systems have evolved from manual digitization with "cursor-clicks" to programming methodologies providing real-time tracking and automized digitizing while offering faster analysis in the process (Winter 2009). In addition, motion analysis systems have become more commercially available, and they are providing very practical way to track human motion in real-time. (Allard et al. 1998, 42-50; Robertson et al. 2014, 13.)

Currently kinematic analysis relies heavily on use of optoelectronic systems consisting of one or multiple cameras and markers attached on the studied body to enable motion analysis. Optoelectronic measurement systems are currently described as the most accurate and golden standard for movement analysis (van der Kruk & Reijne 2018). Novel approaches which do not rely on using markers attached to the body or multi-camera system to produce movement analysis have become more apparent in the last decade. These systems can rely on depth sensor cameras, visual hull and other algorithms to enable motion-analysis.

2.1 Marker-based motion analysis

Marker-based optoelectronic measurement systems are currently the gold standard of pose estimation and motion analysis (van der Kruk & Reijne 2018). It is important to have base knowledge and understanding on methodology of these systems, since they work as point of reference for new systems when their system performance is being validated and the comparisons are made between the two.

The most typical way to collect kinematic data is to use camera or motion-capture system to record location of spherical and reflective markers attached to desired positions on the subject. The output becomes the movement and trajectories of each individual marker in the recording. This is followed by either manual or automatic digitizing to give out the coordinates of each individual marker. After this the coordinates are processed to produce information about segments joints and pose. (Allard et al. 1998; Robertson et al. 2014)

Usually imaging systems use either video, digital video, or charge-coupled device cameras, which record reflection or ambient light of markers. In laboratory setting cameras emit their own light and record movement of the reflective markers on the body, or in some instances active markers that are emitting an infrared light. When studying planar 2-D motion, only one camera placed perpendicular to the plane of motion is necessary, whereas when studying 3-D motion, multi-camera systems are required. By utilizing the cartesian coordinate system, quantification of the position of markers in a recording can ultimately be used to analyse the displacement, velocity, and acceleration. These values can be used to describe segments, joint angles, and angular kinematics, and hence describe the movement of human body in the recording. (Robertson et al. 2014)

Marker-based motion analysis includes very often the use of body models to describe and analyse the pose and the movement as completely as possible. Biomechanical models may have strong anatomical correlation and consist of real joint constraints, while some models are more kinematically oriented and used to reproduce joint kinematics. The most broadly used model for analysing the lower extremities in optoelectronic systems is the plug-in gait model. This model can be expanded to produce full-body plug-in gait model, which is represented as an example in figure 1. The same figure demonstrates simplistically the stages of motion capture with marker-based systems. (Leardini et al. 2017; Klöpfer-Krämer et al. 2020)

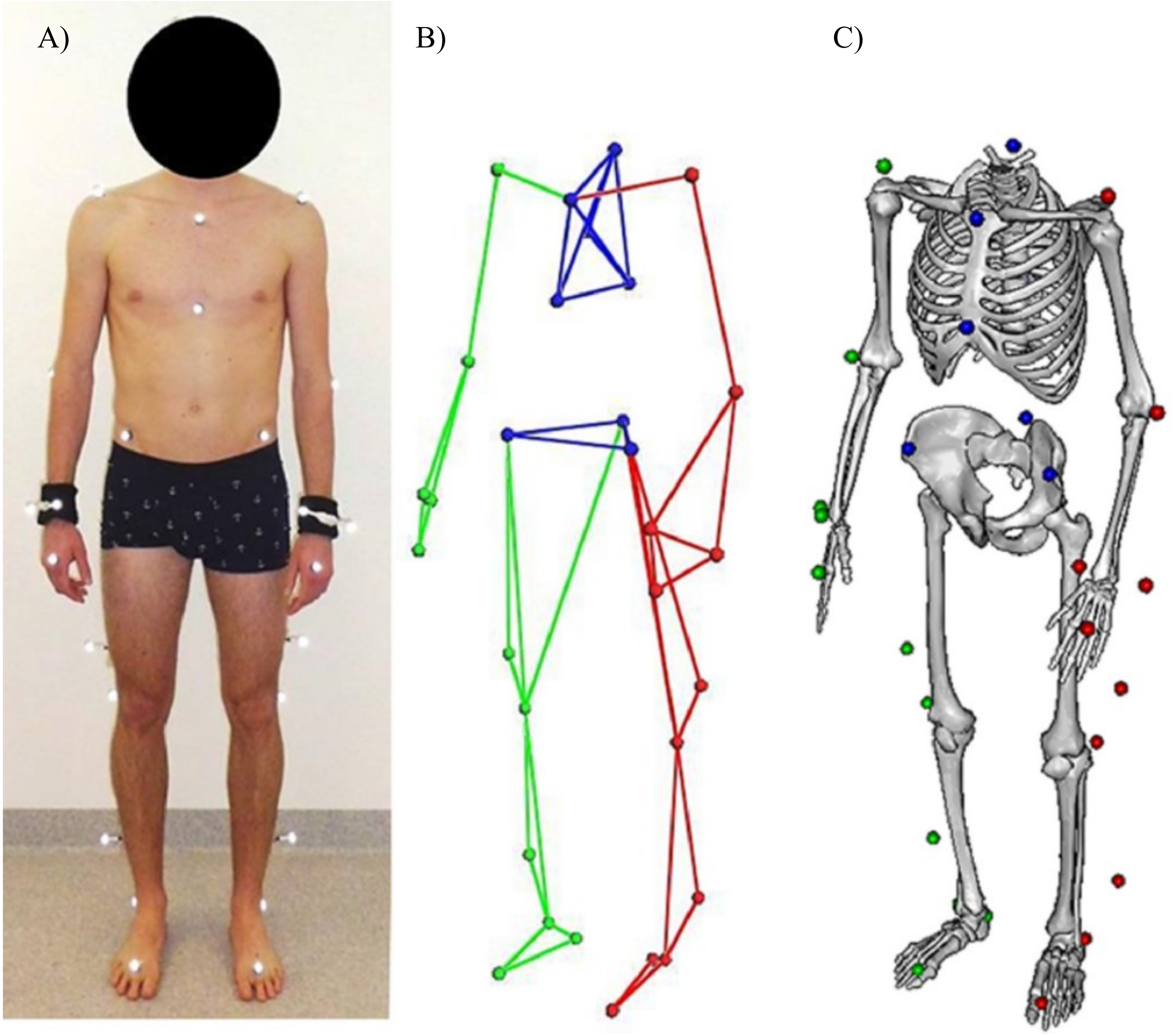

FIGURE 1: Marker-based motion analysis model construction from subject with attached reflective markers (A) to plug-in-gait model (B) to illustration of human skeleton (C). (Klöpfer-Krämer et al. 2020)

2.2 Markerless motion analysis

In some situations, it is impractical to attach markers to the studied subject due to the movement itself, such as competition performance (Robertson et al. 2014). Additionally, using marker-based optoelectronic systems are often expensive and more time consuming than markerless methods and may obstruct the natural movement of a given task (Mathis et al. 2018; Van der Kruk & Reijne, 2018; Colyer et al. 2018). Furthermore, raw video data can be analysed even post-hoc with markerless methods (Mathis et al. 2020). Thus, it would be time-efficient and beneficial to have valid and reliable markerless motion analysis systems for both research and practice. There are few methodological approaches in which this kind of automated markerless motion analysis is currently possible. These methods currently rely on either inertial measurement units (IMUs) or systems utilizing human body models, computer vision and machine learning algorithms (Colyer et al. 2018). Because the theory of IMUs differ substantially from the camera-based systems the focus in the upcoming parts will be on camera-based markerless systems and computer vision.

Vision based markerless systems consist of camera system used, the human body model, image features used, and the algorithms used to determine the body model parameters such as pose and shape. The algorithms are defined as generative or discriminative, where model parameters are compared to the captured image data to define best possible fit or captured image data is used to deduce model parameters, respectively. The general process of forming a pose in markerless systems is depicted in figure 2. As seen in the figure, this process can be divided to offline stage where body models or machine learning based algorithms are introduced, capture stage where image data is captured, processed and put into the algorithms that will form the output of final pose and shape of the body. (Colyer et al. 2018)

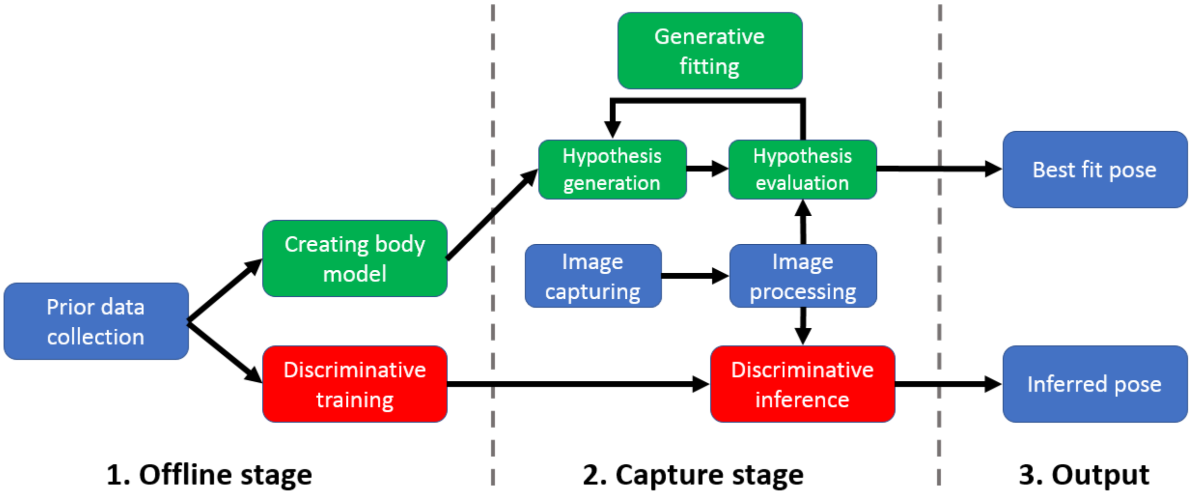

FIGURE 2: General process and stages of forming a pose in markerless motion analysis systems (modified from Colyer et al. 2018).

2.2.1 Image capturing and processing

Camera-systems used for the image capturing can be divided in two groups: depth-map cameras and colour video cameras. Depth-map cameras produce an image, in which every pixel describes the distance of this point from the camera. Depth-map cameras branch further to cameras that have binocular-stereo vision, which basically sense the distance of the object from two different pictures (thus named binocular), and light emitting "active" cameras which sense the distance from the reflection of the area of interest. Active cameras use usually either time-of-flight, in which return time of single light pulse back to the camera is measured, or structured light systems, in which the depth is sensed by distortions of a certain pattern projected on the measurement area. (Colyer et al. 2018)

Active depth sensing cameras which also capture color, such as Kinect, have been shown to be effective tools in interactive applications (Colyer et al. 2018). However, they do not reach the accuracy of traditional motion analysis systems in producing precise biomechanical pose estimation of a sports movement, even though there are implications of equally good reliability and accuracy for a limited number of movement parameters, especially in slower sports movements such as single leg squats and gait and simulated joint movement with a jig (Colyer et al. 2018; Eltoukhy et al. 2016; Kobsar et al. 2019; Mentiplay et al. 2018; Schmitz et al. 2014; Klöpfer-Krämer et al. 2020). Furthermore, application of these systems to sports biomechanics may not be ideal due to their low capture frequencies, ineffectiveness in longer ranges and detrimental effect of direct sunlight to the measurements (Colyer et al. 2018).

Distinction of meaningful image features is subject of importance in computer vision and image processing. There are currently several possibilities in creating a markerless pose, and one of those includes using visual hull, which approximates the subject of interest in the space. Visual hull relies on multiple cameras creating an image silhouette by utilizing chroma keying, where coloring is used to paint the subject with one color and background with another. The combination of these silhouettes from multiple cameras produces the visual hull, and it has shown very accurate results for an automated markerless motion capture system (figure 3). (Colyer et al. 2018)

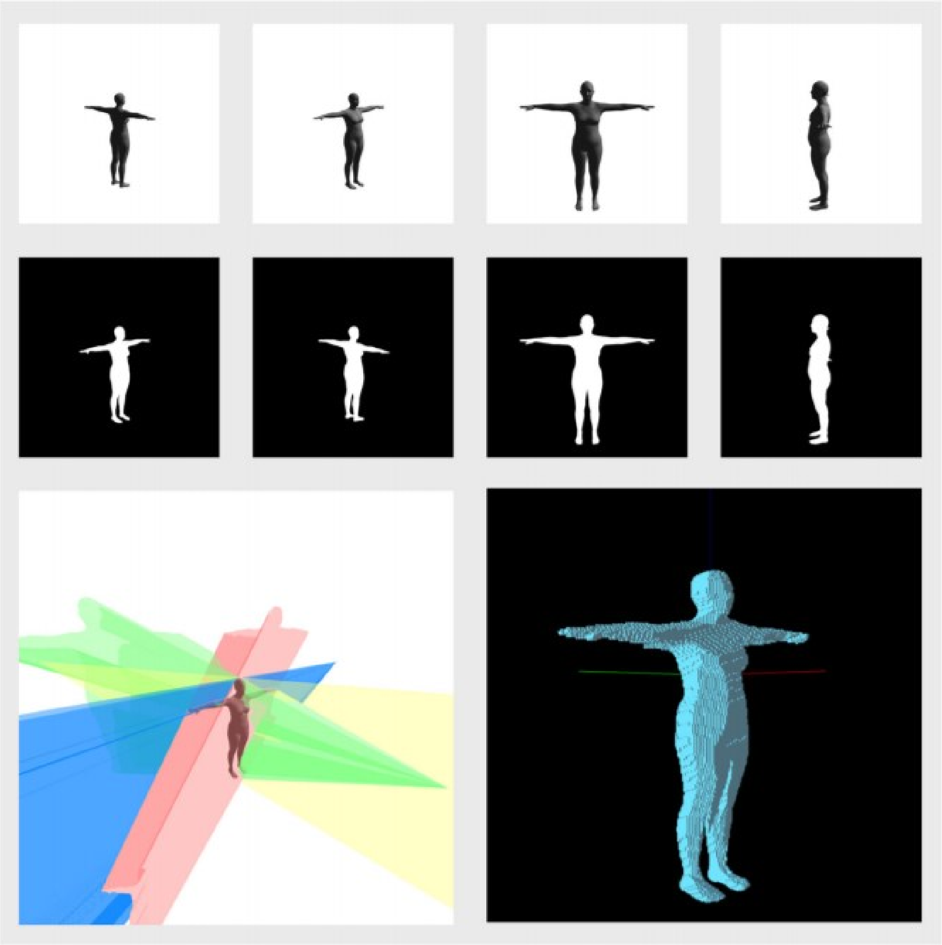

FIGURE 3: The process of 2D image capture (top row pictures), conversion to 2D silhouettes (middle row pictures) and back-projection and generation of visual hull (bottom row pictures). (Colyer et al. 2018)

Another possibility to extract information from images is to utilize pose estimation algorithms. Raw video consists of collection a of images and the pixels they are constructed of. Inside the pictures, the interest of study usually involves objects and their location, scale, and orientation. These objects can be decomposed into multiple keypoints to give semantic meaning to objects such as certain body parts of a human subject. Simultaneously these keypoints provide the given x- and y-coordinates of these points and useful information about the movement of the subject to a researcher or practitioner. (Mathis et al. 2020)

Markerless pose estimation algorithms extract and map this information from images directly. Typically, this requires a set of example images, a training set. In summary, the function of pose estimation algorithms is the mapping of pixels of images to body part coordinates (figure 4). The application and the training data that are used define the labelled body parts that the algorithm returns, which emphasizes the effect of algorithm customization for the functioning and behaviour of the applications. (Mathis et al. 2020)

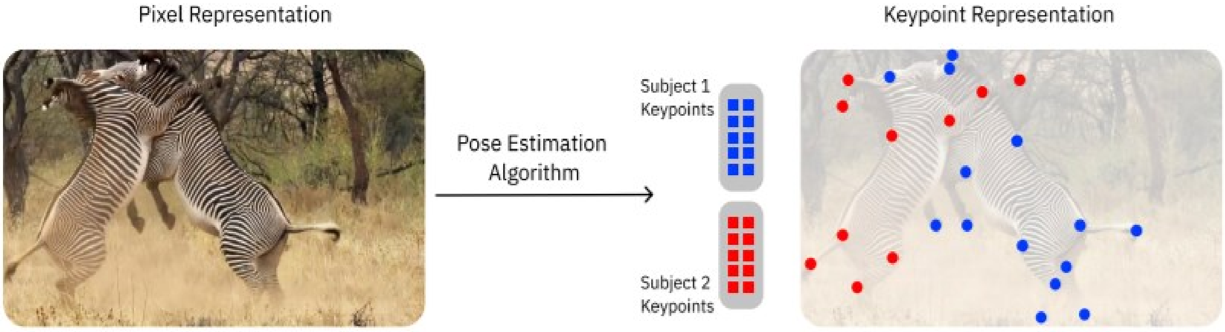

FIGURE 4: Pixel representation and key point representation of an image (Mathis et al. 2020).

2.2.2 Creating body model

The use of kinematic models has many advantages also in markerless motion capture. The body models in markerless motion analysis are similar to those in optoelectronic systems: a biomechanical skeleton is defined as a set of joints and the bones. Their parameters include the length of these bones and the respective joint angles of bones associated with that joint. For discriminative analysis these values are often enough, whereas generative approaches require information about the volume of the subject. (Colyer et al. 2018)

Depending on the system and body models used, models may have strong anatomical correspondence by having real joint constraints, while some models are more kinematically oriented and used to reproduce joint kinematics. In generative automated computer vision, the model representation is usually in the form of “spatial 3D gaussians”, where these spatial gaussians are attached to the skeleton model (figure 5). These generative models may also be produced by high-accuracy 3D scanning or generic statistical 3D models utilizing 3D triangle meshes. (Leardini et al. 2017; Colyer et al. 2018)

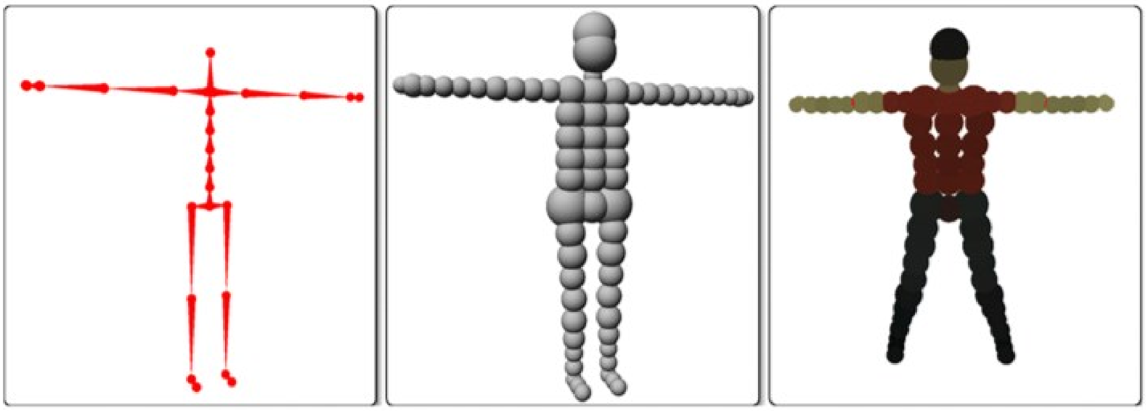

FIGURE 5: Formation of generative 3D spatial gaussian model (middle, right) from simple skeletal model (left). (Colyer et al. 2018)

2.2.3 Generative vs. discriminative approaches

The main function of generative approaches is the comparison of a captured image to a specific hypothesis. Generation of body model is followed by comparison of features from captured images and calculating error value, describing how much the captured image differs from the body model (Colyer et al. 2018). The final output of generative approaches is the best-fit pose, and its name describes this approach appropriately: The best fitting pose of the model is computed according to the provided image in each frame of capture.

In contrast to generative approaches, discriminative approaches do not try to process and fit body model parameters to the image and are also referred to as model-free algorithms. Discriminative algorithms can be divided in two separate methods. The first one includes using image features directly to describe the pose of a subject. This can be done by using machine learning based regressions, where the computer is taught how to determine the pose of the skeleton using image data. Most recently deep learning has been utilized to track particular body parts of multiple people in a supervised way. (Colyer et al. 2018; Mathis et al. 2018)

The second procedure includes a large database of pose examples and subsequently searching for a matching pose from this database to the captured image. Creation of this database can be however quite demanding, since it needs to include an enormous number of pose examples from different camera angles. If there is not enough data, results could easily become invalid. The accuracy is hence dependent on what the system already knows and likely blind to any small new variations that might exist in the captured data. (Colyer et al. 2018)

2.3 Motion analysis with machine learning algorithms

2.3.1 Machine Learning

Machine learning is a rapidly expanding area in computer science, with numerous possible applications. Machine learning is the automated process to detect meaningful patterns in data (Shai & Shai 2014). More practically defined, machine learning is the process in which algorithms learn from given data to build an automated model and perform tasks without explicit programming (Cust et al. 2019). Machine learning requires incorporation of previous knowledge and biasing the learning process for the success of learning algorithms. This means that stronger prior knowledge or assumptions yields an easier learning process. However, stronger prior knowledge also leads consequently to less flexible learning (Shai & Shai 2014). One simple example of machine learning algorithm is called "naïve Bayes", which can, for example, be used to separate spam e-mails from important e-mails (Goodfellow et al. 2016).

It is important to consider the relationship between artificial intelligence (AI) and machine learning. AI describes the use of a computer to model intelligent behaviour with minimal involvement of human control (Kalmet et al. 2020). Thus, machine learning can be seen as AI since it adjusts intelligent learning process into detecting meaningful patterns by parsing information from large datasets (Shai & Shai 2014). Hence machine learning behaves as sub-discipline within AI.

Machine learning systems are composed of four elements: dataset, model, criterion as loss function, and optimization algorithm (Mathis et al. 2020). The input to an automated learning algorithm is certain training data, and the output may take the form of another computer program that can perform a certain task (Shai & Shai 2014). The output in an automated learning algorithm can be described as response function hσ(x), that will predict a ground truth variable y from input vector of variables x. Then models can be introduced to either classification techniques to predict a target class or regressions to predict discrete continuous or discrete values. Ultimately, the models will find an optimal set of parameters to describe the response function and simultaneously make predictions on new and yet unobserved and unlabelled data to minimize a loss function. In sports biomechanics the data inputs can be obtained for example from inertial measurement units (IMUs) or cameras. (Cust et al. 2019)

A very simplified real-life learning example would be picking out fresh and good tasting oranges from a big bowl with raw, fresh, and rotten oranges. Here, oranges represent the input, variables x would describe the properties of an orange, such as colour, hardness, and a particular smell, and variable y (the label) would describe the taste of the orange. Then a model could be introduced to select oranges that taste good (i.e. have bright orange colour and have certain hardness and smell) by utilizing a loss function (i.e. mapping out good tasting oranges from bad tasting oranges by giving out penalty for not achieving good taste). Now the model could describe the optimal set of parameters that yield a tasteful orange. These responses would create predictions for new patches of oranges and the learning process would end to picking even better tasting oranges according to certain combination of parameters x.

Model training can be altered by using either supervised learning, unsupervised learning, or semi-supervised learning (Cust et al. 2019). Supervised learning includes introducing certain significant information (a label) to a learning algorithm to direct a faster and more precise learning process. There is an environmental factor that teaches and supervises the learner by giving additional information as labels (Shai & Shai 2014). In the previous orange example, such label could be “rotten oranges”, describing oranges that can be excluded right away. When looking at unsupervised learning however, there is no clear distinction between training and test data. Learner processes the input data to produce a summary or compressed version of a given data set, for example as clusters of similar objects. Semi-supervised learning or intermediate learning setting provides some additional information to the learner, but it is somehow limited and requires the learner to predict more information for test examples than what was available at learning setting (Shai & Shai 2014). In the previous orange example, this could be explained in such way that there is only some vague label available as information, such as “face expression” of the person who tastes the orange. This would only create more predictions to learning due to the many possible reasons why a person has a certain face expression (for example mood and familiarity with eating oranges), rather than qualifying oranges based on their own qualities.

When using images as databases feature representations are a necessity to recognize complex object in the images (Goodfellow et al. 2016; Mathis et al. 2020). In this process the whole image representation (i.e. surroundings) is recognized together with mapped features. This learning mechanism is very beneficial, since human processing of complex images is much more laborious, slower and less accurate. However, since numerous factors cause variation and affect how an object is seen in the image (i.e. lighting, camera-angle, shadows, reflections etc.), it can be hard to disentangle and remove unnecessary factors of variation. Deep learning provides a great possibility to fix this issue by expressing the representations relative to other simpler representations. In other words, deep learning allows a computer model to gradually learn more and more abstract features in order to recognise objects. (Goodfellow et al. 2016)

2.3.2 Deep learning

Deep learning, or in other words artificial neural networks, is one branch of machine learning, where deep neural network is inspired from the architecture of biological neural networks of the human brain. These hierarchical models create a deep architecture consisting of representative learning hidden layers. When compared to machine learning, the advantage is that these computational learning models allow data input features to be automatically extracted from raw data and transformed to handle unstructured data (Cust et al. 2019; Goodfellow et al. 2016). Figure 6 demonstrates the basic principles of how deep learning can represent a person as simpler concepts in an image (Goodfellow et al. 2016). As seen in the figure, deep learning works in the multilayer perception, or feedforward deep network. As in machine learning, each layer works as an input-output function, where each output provides a representation as input for each new layer. In the figure, the input is represented as observable visible layer. Now, as the computer cannot extract the wanted features (meaning the definition of a person in the image in this example) directly from the pixels of the image, they are extracted on more abstract hidden layers. The edges of pixels can be identified easily from the picture on the first hidden layer by comparing the brightness of adjacent pixels. The second layer can be deducted to describe the corners and contours, which are identifiable as certain compilation of edges. From here the third layer can be used to detect different parts of particular objects by finding certain pattern of corners and contours. From here the specific combination of these object parts can be used to describe and recognize a specific object in the image as the final output.

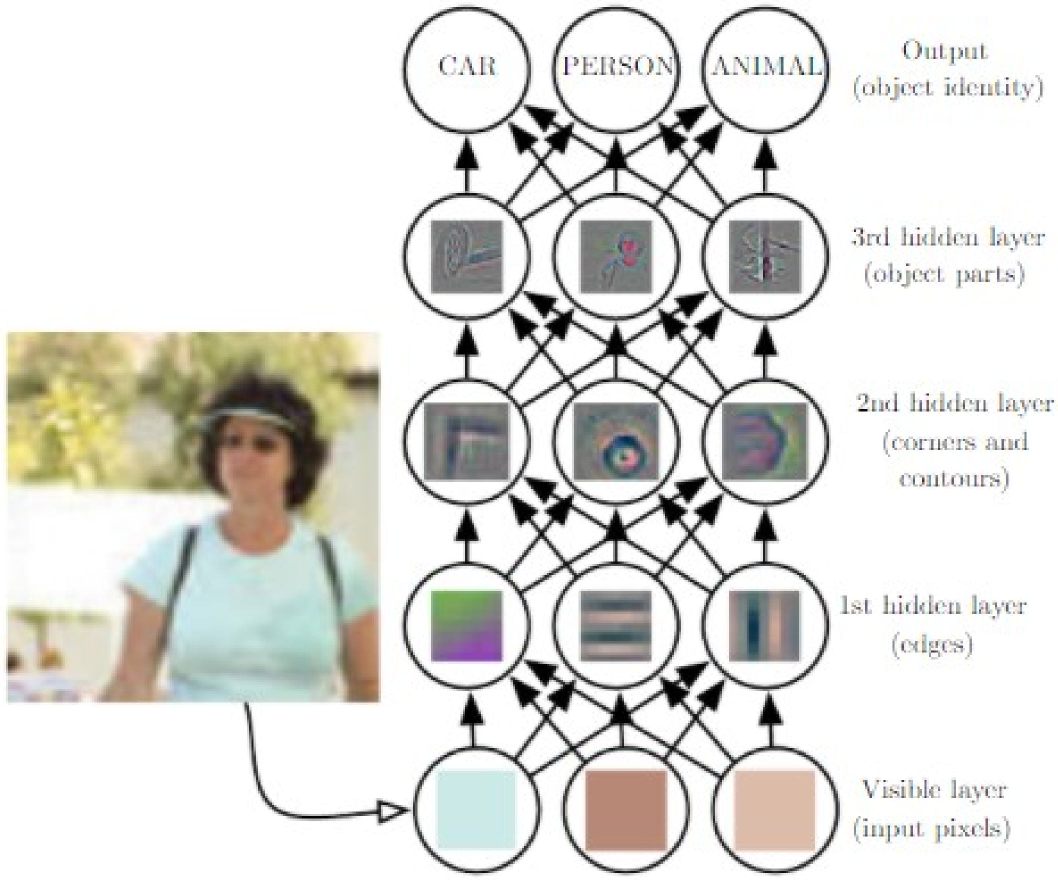

FIGURE 6: Example of deep learning model. (Goodfellow et al. 2016)

2.3.3 Pose estimation

Pose estimation falls directly under object detection by feedforward deep network. Deep learning algorithms have even been benchmarked as the best algorithms to create human pose estimations. In pose estimation from an image, the working system can be viewed consisting of an “encoder”, which extracts features from the images, and “decoder” creating body part location predictions from these features, similarly to previous example. Encoders and decoders are also known as the backbone and output heads, respectively. (Mathis et al. 2020)

Currently, encoders, and decoders work as deep neural networks (DNN) optimized on the pose estimation. The ideal process for estimating a pose is to learn representations from raw video or image datasets (encoding part) and creating predictive model for the human pose (decoding part). Practically this can be produced with a sequence of differentiable and non-linear transformative multiple layers, and by utilizing back-propagation algorithm for the model as a whole (Mathis et al. 2020). As opposed to forward propagation by feedforward neural network, where an input x levels up to next hidden layer to produce output y, continuing until final output, back-propagation allows final output information to flow back to determine the gradient of the function in hand (Goodfellow et al. 2016). It is worth mentioning that according to Mathis et al. (2020) DNN-based tools optimize the feature representation exactly for pose estimation tasks.

The input-output relationships that a model should learn are defined by the datasets. When performing pose estimation, the output is defined as a specific pose, and the input is defined a specific image (Figure 7). Composed body model will be updated by the optimisation algorithm to reduce the size of the loss function. Here the goal of the loss function is to measure the similarity between the predicted and the ground truth value. Properties of these system elements will define how the pose-estimation system works and behaves. (Mathis et al. 2020)

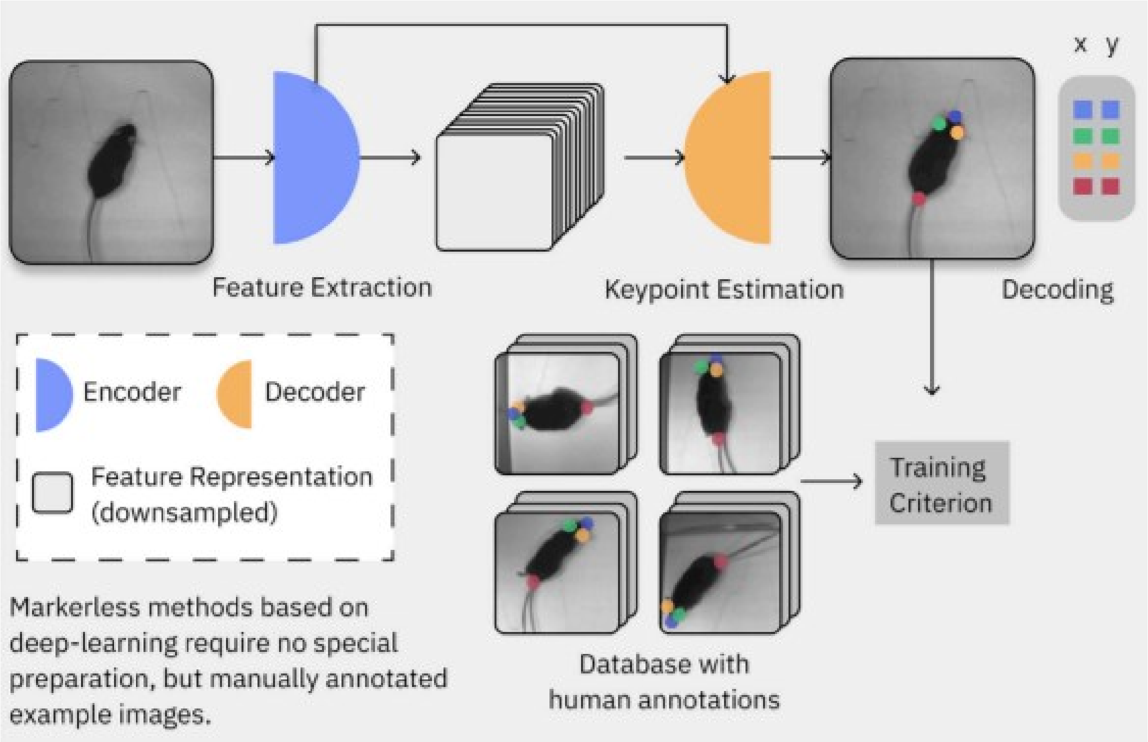
x y
Feature Extraction
Keypoint Estimation
Decoding
Encoder
Decoder
Feature Representation (downsampled)
Training Criterion
Database with human annotations
Markerless methods based on deep-learning require no special preparation, but manually annotated example images.

FIGURE 7: Machine learning model training by using human annotated key points. (Mathis et al. 2020)

As discussed earlier, discriminative methods can be divided in two depending on the datasets: the data can be extracted from one or multiple datasets or from self-captured raw image or video data. These two are also essential when discussing deep learning as a tool for pose estimation. Earlier gathered databases can be used for pretraining of computer vision models. The datasets considering image recognition are commonly much larger than those datasets for pose estimation purposes: Image recognition datasets such as ImageNet consists of over 14,2 million images in 21 000 different classes, whereas benchmark pose estimation datasets such as MPII pose consist of 40 000 images of 26 000 individuals. The computer can be pre-trained with these datasets by using transfer learning. Transfer learning is described as the ability to use parameters from a network trained for a certain task as part of another network to enhance its performance. (Mathis et al. 2020) An example of this could be a model trained to detect object classes (for example a dog, a cat or a table) and re-train it to perform pose estimation.

When using self-captured image or video data, relatively small number of images are sufficient for model training, especially when the data is captured in laboratory setting. Datasets used sufficient to create a model for task of interest commonly contain 200 to 500 images (Mathis et al. 2018; Cronin et al. 2019). Cronin et al. (2019) demonstrated that 300 to 400 images are adequate to train a neural network to recognize and label body parts as accurately as manually performed labelling by a human. However, when it is possible, Mathis et al. (2020) recommend using pre-trained pose estimation algorithms since they can save time, increase robustness and less training data is required. Also, using encoder architectures that have been pre-trained on larger scale datasets has been demonstrated to be beneficial for pose estimation measurements for relatively small laboratory setting (Mathis et al. 2018; Mathis et al. 2020).

A model in deep learning describes the statistical operations that are involved in the development of an automated prediction (Cust et al. 2019). Deep learning models can use generic encoder architectures usually based on object recognition. There are several different architecture designs that are used for pose estimation. The use of these architectures is advantageous since they affect the most critical properties of the algorithms. These include for example training-data requirements as mentioned earlier, algorithm inference speed, and memory demands of the computer. An example of commonly used backbone architecture is residual networks (ResNets). Convolutional neural networks (CNN) are shallow backbone architectures and are defined as DNN consisting of multiple convolutional layers. Their optimization becomes hard and performance decreases when the number of layers increases over 20. ResNets provide a possibility to networks with much larger depths without weakening the output. Their working principle differs from CNNs by adding the input value to the loss function ($y = x + f(x)$, instead of $y = f(x)$). ResNets provide improved optimization and regularizes the loss function. (Mathis et al. 2020)

Keypoints of desired body parts on the object can be represented as simple coordinates on image data. Deep learning utilizes loss function to determine the loci of these points, and there are two methods to achieve this. The keypoints can be deducted by using regression or more commonly, by using grid same size as the input image to create a heatmap of location probabilities for every body part (figure 8). (Mathis et al. 2020)

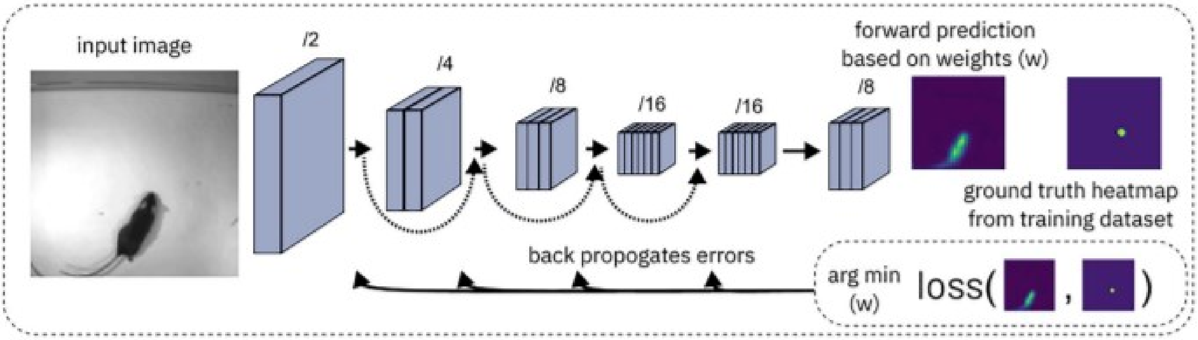

FIGURE 8: One possibility for the network training with heatmaps. Input image is processed through feedforward network, and the target heatmap is compared to the forward prediction heatmap. Now with back propagating the loss (which measures the difference between predicted and target or ground truth heatmap) can be minimized and the network parameters are optimized. (Mathis et al. 2020)

Most pose estimation packages work using the same principles. Currently there are about 10 packages available for use, each with different focus on certain user experience, networks and efficiency and accuracy of handling the data. Currently one of the most cited and used deep learning tool for animal and human motion capture is DeepLabCut. When using these deep learning tools, possible errors may occur due to low video quality and poor labelling quality and quantity. (Mathis et al. 2020) This emphasises the importance of labelling accuracy and proper set-up for measurements to ensure accurate pose estimation by using deep learning.

2.3.4 Accuracy

When discussing about accuracy of pose estimation by deep learning, one way to approach it is to compare the performance of human labeller and the computed pose estimation. Cronin et al. (2019) evaluated deep neural network performance (DeepLabCut) by comparing pairwise Euclidean distances of marker locations with root mean square error in underwater running. When comparing a training set trained with 500 images, the results showed an error of 2,92 pixels, or about 1 centimetre. These represented similar values over models trained over 400 and 300 images. The root mean square error started to increase significantly when the models were trained with less than 300 images (figure 9). These results imply that using 300 images or more for training the model is sufficient for the same accuracy as a human labeller (Cronin et al. 2019), although this depends on the dataset used.

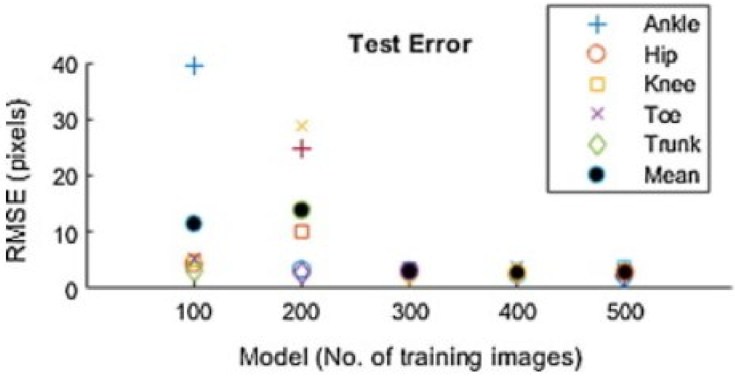

FIGURE 9: Root mean square error (RMSE) increases when using less than 300 images for training a model. Using 300 images or more can be sufficient to reach as accurate pose estimation as a human labeller would give.

2.3.5 Validity

There have been few studies comparing deep neural networks performance to marker-based systems. Strong agreement has been demonstrated between markerless and marker-based motion capture for sagittal plane kinematics in vertical jump with DeepLabCut and bilateral squat with OpenPose (Drazan et al. 2021; Ota et al. 2020). Ota et al. (2021) demonstrated similar results with good to excellent values of intraclass correlation in sagittal kinematics of walking and running on a treadmill by comparing the performance of a markerless system (OpenPose) to an optoelectronic system (Vicon). However, for frontal plane kinematics, poor intraclass correlation values and no significant correlations were found in walking and running on a treadmill between these systems (Ota et al. 2021). Thus, there seems to be good level of validity for markerless systems in sagittal plane kinematics, but not yet so in frontal plane kinematics.

2.3.6 Repeatability

There does not seem to be any studies published yet considering the repeatability of markerless kinematic analysis systems. If the accuracy and validity of these novel systems prove to be reliable, more studies are necessary to determine whether these systems yield reliable test-retest values. In comparison, multiple studies demonstrate that the repeatability of optoelectronic systems appears to be reliable in walking and running activities, with within-session measurements showing a higher test-retest reliability compared to measurements separated by at least two days (Wright et al. 2011; Sinclair et al. 2012; Leszczewska et al. 2012; Judson et al. 2020).

It is important to note that when assessing the repeatability of motion analysis systems, the measured movement itself may affect the result. Wren et al. (2020) demonstrated that there exists a high trial-to-trial variability in the joint range of motion parameters when performing different sports tasks. This highlights that the end results in test-retest reliability measurements may be dependent on the assessed sports movement.

3 RESEARCH QUESTIONS AND HYPOTHESES

As deep learning approach is becoming more prominent in kinematic analysis, and as there are not yet published studies about its repeatability, which is one of the important factors defining the validity of a measurement system, there is an increasing demand for researching this area of topic. Hence, the aim of this study was to evaluate kinematic analysis repeatability by a deep learning approach and its repeatability in consecutive countermovement jumps (within-day and between-day) for athletes. The interest focused on calculating the joint angles for each jump trial, and then estimating intraclass correlation coefficients for each subject’s first and second trial. Finally, the aim was to use this correlation data to calculate the mean and standard deviation of these correlations for every angle measured to deduce the test-retest reliability of the deep learning approach. The hypotheses of this study were that the intraclass correlation coefficient for angle data would have low variability between subjects’ trials demonstrating high correlations, and good repeatability for both sagittal and frontal plane angle data.

4 METHODOLOGY

This study included a total of 70 subjects (age: 18.8 ± 2.9 years, height: 177.2 ± 11.3 cm, mass: 74.2 ± 13.8 kg; 39 females and 31 males). The subjects were national elite level recreational athletes in football (38 females), basketball (16 males), ice hockey (9 males), track and field athletes (6 males) and one female team gym athlete. Participants provided written informed consent before participating in the tests and gave information about their injury background. All the test subjects were healthy, non-injured athletes who were familiar with the counter movement jump test. This study was part of the Training Room Project by the Research institute of Olympic Sport. The study received ethical approval from the ethical committee of the University of Jyväskylä and all the tests were conducted according to the Helsinki declaration.

4.1 Measurements

The measurements began with instructed warm-up for each test subject. Guided warm-up consisted of 5 minutes on a cycle ergometer at the subject's preferred pace, followed by dynamic activation movements and stretches. After the warm-up, the test subjects were guided to stand over an "x" marked on the ground and instructed to perform two maximal counter movement jumps, with brief a pause (3-10 seconds) in between each jump. Subjects were instructed to "jump as high as possible with their hands placed on the side of their body". Each performance was recorded with two GoPro 3 hero -cameras (GoPro Inc., San Mateo, CA, US) with a 120 Hz frame rate and resolution of 1280 width and 720 height in pixels. The cameras were placed on two camera-stands at hip height perpendicular to each other, five meters from the measurement area. The first camera faced the subject (frontal view), and the second one faced subject's left side (sagittal view). To synchronise the videos from both cameras for subsequent analysis, a small piece of tape was attached to a stand visible to both cameras and a laser pointer was used to point a laser at the tape at the beginning of each trial. The tests were performed in two separate areas, with different lighting and background. The test set-up is represented in figure 10. The within-day jumps were performed during the same test session with a brief pause between the jumps. Between-day jumps were performed in two different test sessions separated by two weeks.

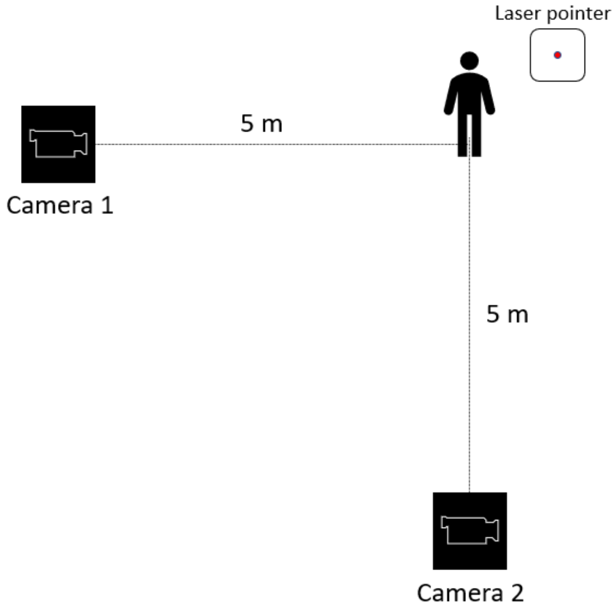

FIGURE 10: The test set-up.

4.2 Computing and training of data

After collection of data, the video files were first synchronised and cropped with an open-source program, Kinovea. The synchronisation was achieved by selecting a frame where the laser was visible for both cameras at the same location on the tape. The jump height was calculated with equation: $\frac{1}{2} \times g \times (t / 2)^2$, where $g = 9{,}81\ \mathrm{m/s^2}$ and t is the flight time. Flight time was defined as the time between the first point when the feet left the ground completely and the point where they touched the ground after the jump. The used application for training the deep learning model was DeepLabCut (Mathis et al. 2018). Individual models were trained for each camera view separately. Both frontal and sagittal models of CMJ performance were each processed by 500 randomly selected test images from 50 subjects (10 images from each) to provide sufficient accuracy in the deep learning process (Cronin et al. 2019). For frontal view videos, the marker locations were labelled for both sides of the body to shoulder (acromion), hip joint (greater trochanter), knee joint (mid-point of patella), ankle joint (mid-point between malleoli) and toes (head of shoe), as seen on figure 11. The marker locations of shoulder (acromion), hip joint (greater trochanter), knee joint (lateral femoral condyle), ankle joint (lateral malleolus) and toes (head of shoe) were manually labelled for sagittal test images, as seen on figure 12. In most cases, all the markers were visible throughout CMJ performance.

These images were then used to train a deep learning network to predict the location of markers and joint angles separately for the frontal and sagittal view camera. The models were trained using Resnet-101 model. Evaluation was performed for remaining 20 subjects (11 for within-day and 9 for between-day evaluation). The jumps from these subjects were not part of creating the model. For the sagittal videos, hip, knee and ankle joint angles were calculated by using atan2 function in Matlab. For the frontal view videos, inspection of knee and ankle angles were selected for the analysis due to the research interest and importance and were similarly deduced with atan2 function. To correct for misplaced or missing markers, raw data was filtered with a median filter and subsequently with Butterworth 4th order low-pass filter.

After filtering, data was further processed with Matlab by first aligning the curve data of consecutive (trial 1 and trial 2) jumps. This was accomplished by inspecting sagittal knee angle: The first point where knee angle increased again after the eccentric and concentric phase of CMJ was deduced for both jump trials. This point would indicate the end of knee extension approximately at take-off. Next the data was cropped accordingly. Starting point was defined as the point where the subject started lowering their centre of mass with the eccentric phase of CMJ. The ending point was defined as the point after landing the CMJ where the subject regained their normal stance. Again, by using the sagittal angle data, the first point where the knee angle increased by more than 5 degrees from average knee angle of standing still before the jump was deduced for each jump. Starting and ending points of cropping were determined by this angle for each jump. If there were differences in knee angle during standing before the CMJ, the jump with larger initial knee angle (meaning lower stance) was used for both jumps to determine the starting and ending points for cropping. This data was then compiled to a table for statistical analysis.

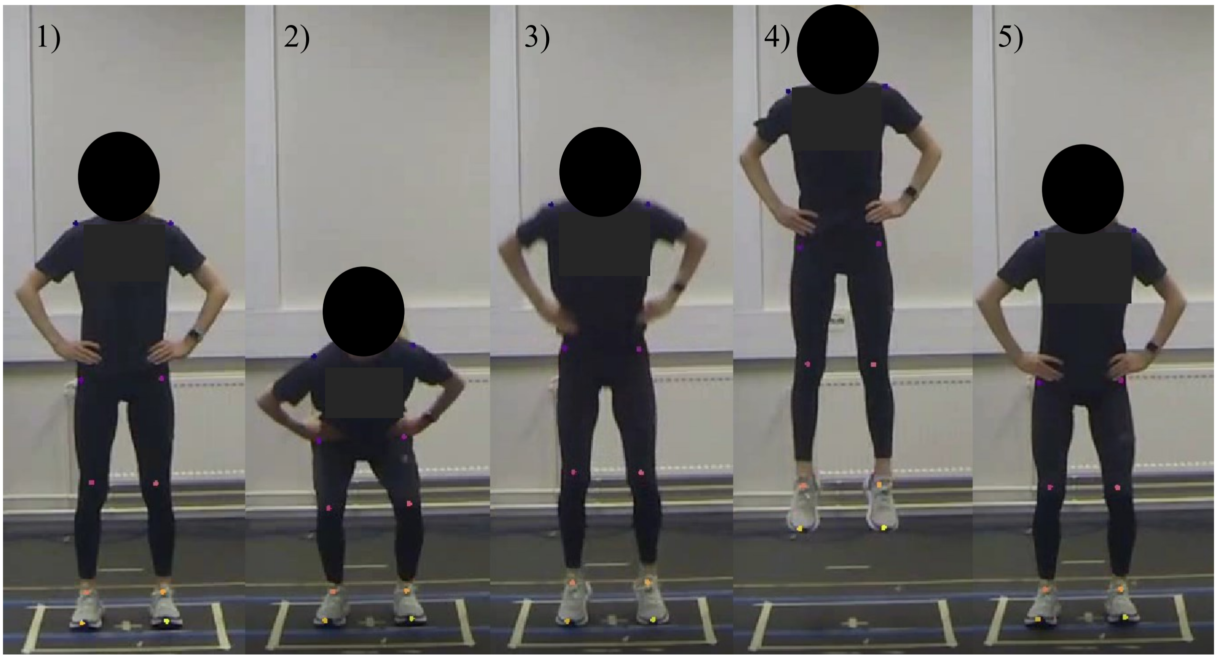

FIGURE 11: Marker locations for frontal view videos, with the locations at shoulder (blue dots), hip joint (purple dots), knee joint (pink dots), ankle joint (orange dots) and at head of the shoe (yellow dots). The images represent the phases of CMJ: 1) initial stance, 2) the end of eccentric phase (the lowest point of center of mass), 3) end of concentric phase (take-off), 4) highest point during the flight, 5) the end of eccentric phase of landing (the lowest point of center of mass at landing).

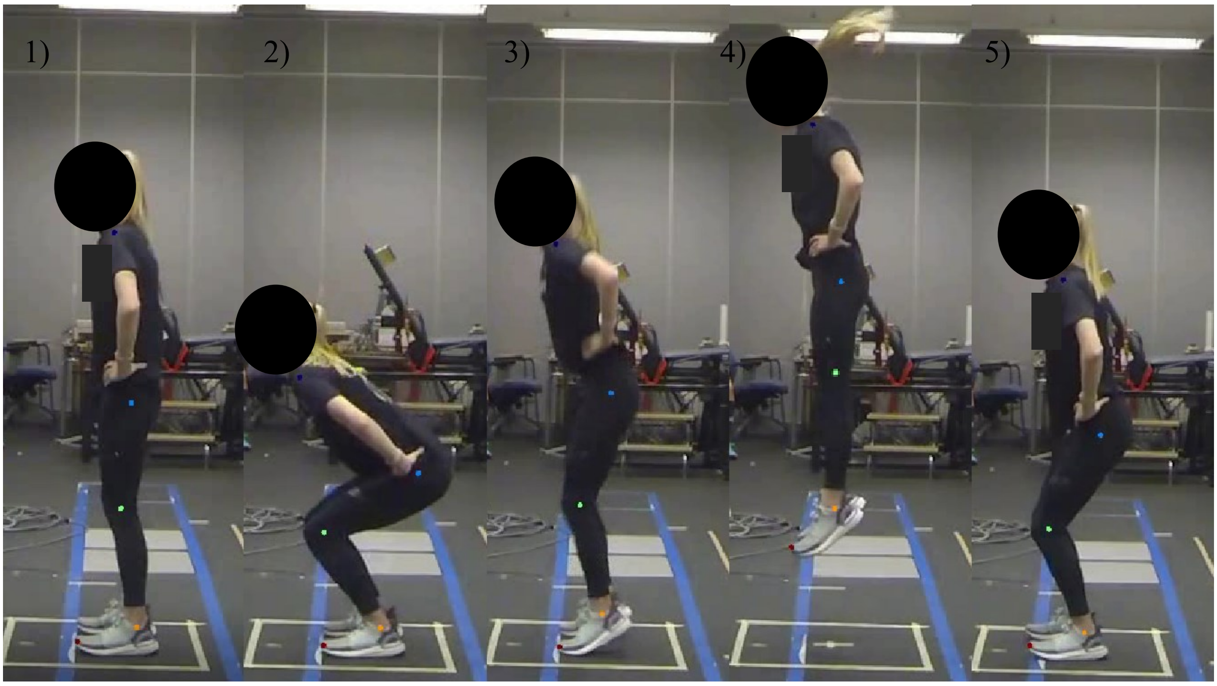

FIGURE 12: Marker locations for sagittal videos, with the locations at shoulder (dark blue dots), hip joint (light blue dots), knee joint (green dots), ankle joint (orange dots) and at head of the shoe (red dots). The images represent the phases of CMJ: 1) initial stance, 2) the end of eccentric phase (the lowest point of center of mass), 3) end of concentric phase (take-off), 4) highest point during the flight, 5) the end of eccentric phase of landing (the lowest point of center of mass at landing). Note that the image representing the take-off (3) is taken one frame before the ground contact is lost and the highest knee extension is achieved.

4.3 Statistical analysis

After all the raw data of marker locations were filtered and exported, the values were imported to IBM SPSS Statistic 26 (IBM Corp., Armonk, NY, US) for statistical analysis. The mean and standard deviation were calculated for jump heights, and test-retest values were calculated by intraclass correlation coefficients (ICC) for subjects whose jumps were not used in the creation of the deep learning model. The ICC model used for test-retest was single measurement two-way mixed effects with absolute agreement, as Koo and Li (2016) recommend in their guideline for selecting and reporting ICC. By using 95 % confidence interval the following values were used to determine the indication of test-retest reliability: Poor (0,5 or less), moderate (0,5 – 0,75), good (0,75 – 0,9) and excellent reliability (0,9 or greater) (Koo & Li 2016).

5 RESULTS

The mean jump height of the first and subsequent countermovement jumps for within-day subjects were 0.30 ± 0.04 m and 0.30 ± 0.03 m, respectively. For between-day subjects the mean jump height was 0.44 ± 0.04 m for the first measurement and 0.43 ± 0.05 m after two-week time. The jump height ICC values were 0.74 and 0.94 for within- and between-day subjects, respectively.

5.1 Sagittal angles and correlations

Sagittal CMJ joint angle data from a typical subject is represented in figure 13. The within-day ICC values for this subject were 0.99, 0.98 and 0.97 for hip, knee and ankle angle, respectively. For within-day subjects (T1–T11), ICC values ranged between 0.87 – 0.99, 0.90 – 0.99 and 0.82 – 0.99 for hip, knee and ankle angles, respectively. For between-day subjects (T12–T20), these values ranged between 0.88 – 0.98, 0.76 – 0.98 and 0.73 – 0.97 for hip, knee and ankle angles, respectively.

The grouped mean ICC values equalled 0.95 ± 0.04, 0.96 ± 0.03 and 0.95 ± 0.05 for within-day subjects, and 0.95 ± 0.03, 0.95 ± 0.07 and 0.89 ± 0.08 for between-day subjects, for hip, knee and ankle angles, respectively. All ICC values and their 95 % confidence intervals for each subject and the grouped means for both within-day and between-day measurement values are displayed in table 1 and table 2, respectively.

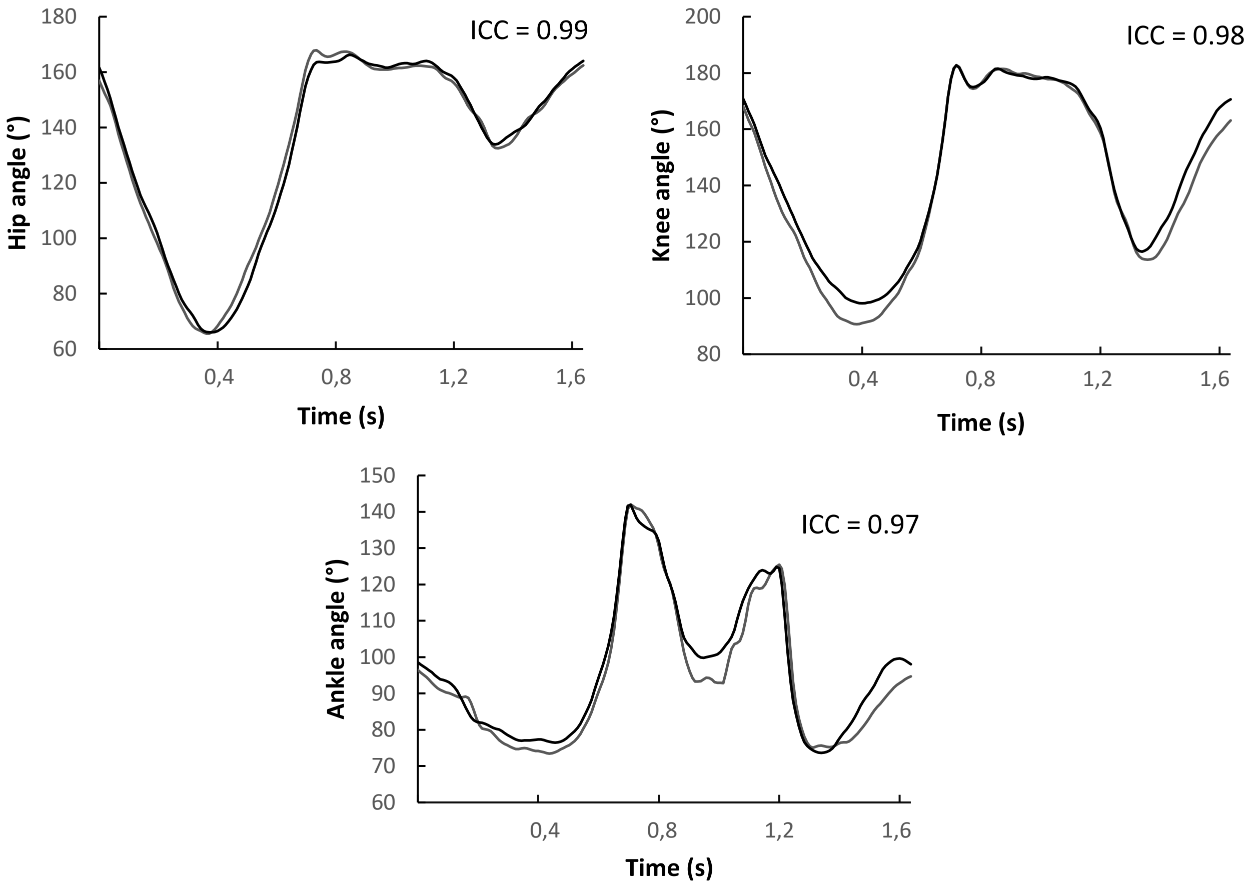

FIGURE 13: Representation of within-day sagittal joint angle data for a single subject. Full extension of each joint is defined at 180° joint angle and decrease in joint angle denotes joint flexion. Hip, knee and ankle joint angles all show high ICC values and thus represents excellent test-retest reliability for this particular subject.

Subject	Hip ICC	CI 95%	Knee ICC	CI 95%	Ankle ICC	CI 95%
T1	0.96	0.95 – 0.97	0.99	0.97 – 0.99	0.94	0.92 – 0.95
T2	0.97	0.96 – 0.98	0.97	0.62 – 0.99	0.93	0.87 – 0.95
T3	0.99	0.99 – 0.99	0.99	0.98 – 0.99	0.97	0.96 – 0.97
T4	0.87	0.55 – 0.94	0.90	0.81 – 0.94	0.82	0.78 – 0.86
T5	0.99	0.99 – 0.99	0.98	0.88 – 0.99	0.97	0.91 – 0.98
T6	0.97	0.97 – 0.98	0.98	0.98 – 0.99	0.90	0.86 – 0.93
T7	0.98	0.96 – 0.98	0.99	0.98 – 0.99	0.99	0.98 – 0.99
T8	0.98	0.95 – 0.99	0.95	0.80 – 0.98	0.96	0.94 – 0.97
T9	0.94	0.92 – 0.95	0.96	0.95 – 0.97	0.98	0.98 – 0.98
T10	0.88	0.61 – 0.94	0.90	0.76 – 0.95	0.98	0.95 – 0.99
T11	0.92	0.87 – 0.95	0.94	0.91 – 0.95	0.98	0.97 – 0.98
Mean	**0.95 ± 0.04**		**0.96 ± 0.03**		**0.95 ± 0.05**	

TABLE 1: ICC values for all the sagittal joint angles of within-day subjects and their grouped means and standard deviations.

Subject	Hip ICC	CI 95%	Knee ICC	CI 95%	Ankle ICC	CI 95%
T12	0.88	0.85 – 0.91	0.76	0.70 – 0.81	0.83	0.64 – 0.91
T13	0.94	0.92 – 0.95	0.98	0.88 – 0.99	0.93	0.70 – 0.97
T14	0.96	0.93 – 0.97	0.95	0.94 – 0.96	0.77	0.70 – 0.83
T15	0.98	0.93 – 0.99	0.96	0.86 – 0.98	0.87	0.83 – 0.90
T16	0.89	0.75 – 0.94	0.95	0.91 – 0.96	0.73	0.59 – 0.81
T17	0.95	0.94 – 0.96	0.97	0.90 – 0.98	0.97	0.91 – 0.98
T18	0.96	0.92 – 0.97	0.96	0.93 – 0.98	0.95	0.93 – 0.96
T19	0.98	0.97 – 0.98	0.97	0.95 – 0.98	0.94	0.92 – 0.95
T20	0.96	0.93 – 0.98	0.98	0.98 – 0.98	0.97	0.96 – 0.98
Mean	**0.95 ± 0.03**		**0.95 ± 0.07**		**0.89 ± 0.08**	

TABLE 2: ICC values for all the sagittal joint angles of between-day subjects and their grouped means and standard deviations.

5.2 Frontal plane angles and correlations

Frontal plane CMJ joint angle data from a typical subject is represented in figure 14. The within-day ICC values for this subject were 0.64, 0.97, 0.50 and 0.73 for right knee valgus, left knee valgus, right ankle and left ankle angle, respectively. For within-day subjects, ICC values ranged between -0.14 – 0.93, 0.31 – 0.97, 0.19 – 0.91 and -0.03 – 0.84 for right knee, left knee, right ankle and left ankle angles, respectively. For between-day subjects, these values ranged between 0.53 – 0.88, 0.13 – 0.82, 0.22 – 0.77 and 0.04 – 0.74 for right knee, left knee, right ankle and left ankle angles, respectively.

The grouped mean ICC values equalled 0.43 ± 0.31 for right knee angle, 0.68 ± 0.23 for left knee angle, 0.62 ± 0.22 for right ankle angle and 0.53 ± 0.29 for left ankle angle in within-day subjects. For between-day subjects, ICC values were 0.75 ± 0.10 for right knee angle, 0.49 ± 0.27 left knee angle, 0.53 ± 0.17 for right ankle angle and 0.34 ± 0.26 for left ankle angle. All ICC values and their 95 % confidence intervals for each subject and the grouped means for within-day subjects are displayed in table 3 and table 4, and for between-day subjects they are displayed in table 5 and table 6.

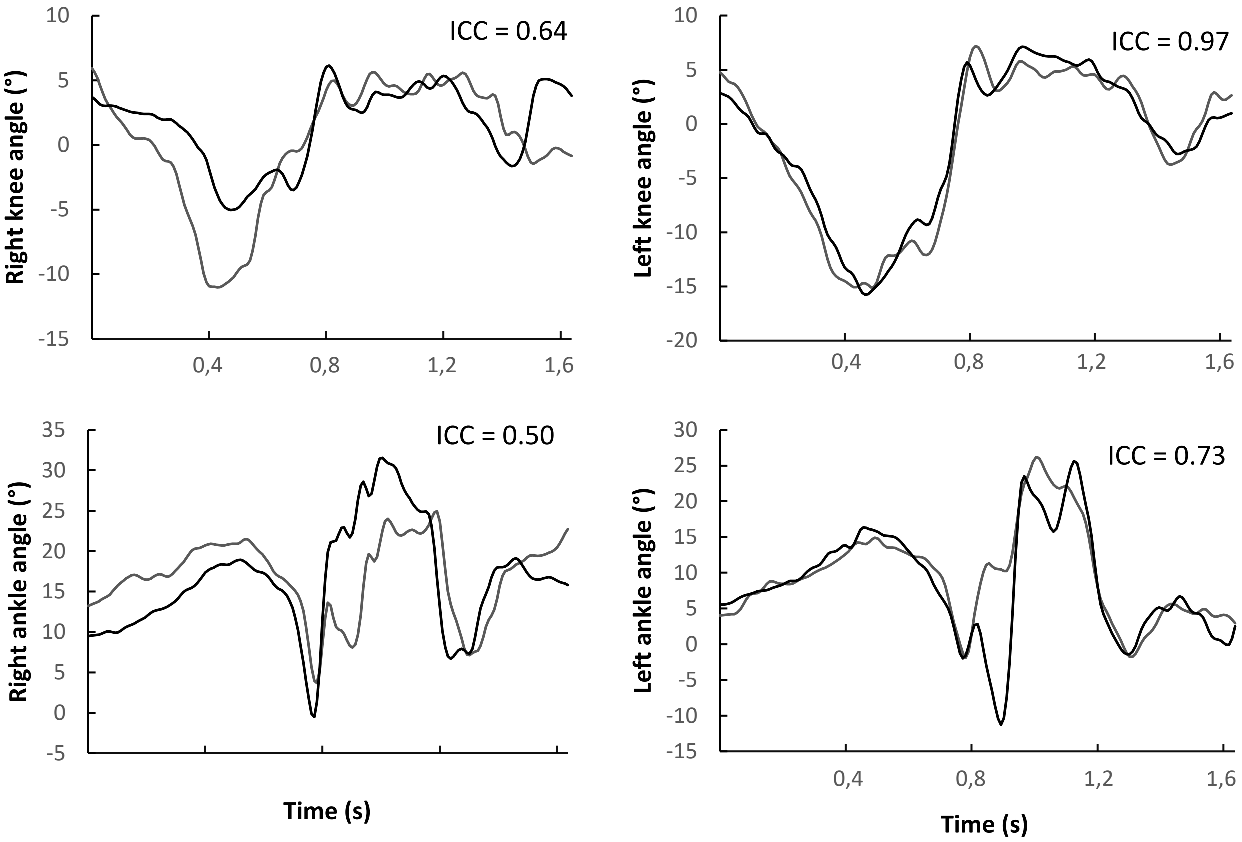

FIGURE 14: Representation of frontal plane joint angle data for a single subject. For knee joint, positive values represent knee valgus and negative values knee varus. For ankle joint, positive values may represent ankle pronation and external rotation at knee and hip joint, and negative values may represent ankle supination and internal rotation at knee and hip joint when the sole of the foot is in contact with the ground. Zero angle defines naturally the neutral position of the joint. For this subject, left knee angle represents excellent test-retest reliability, whereas right knee, and both ankle angles demonstrate only moderate test-retest reliability.

Subject	Right Knee ICC	CI 95%	Left Knee ICC	CI 95%
T1	0.64	0.48 – 0.74	0.97	0.96 – 0.98
T2	0.53	0.19 – 0.71	0.31	0.17 – 0.44
T3	0.45	0.32 – 0.55	0.90	0.78 – 0.94
T4	0.40	0.26 – 0.51	0.48	0.38 – 0.57
T5	-0.06	-0.18 – 0.06	0.94	0.92 – 0.95
T6	0.93	0.91 – 0.95	0.89	0.85 – 0.91
T7	0.47	0.24 – 0.63	0.68	0.50 – 0.79
T8	0.67	0.25 – 0.83	0.81	0.75 – 0.86
T9	0.64	0.45 – 0.75	0.55	0.44 – 0.64
T10	-0.05	-0.16 – 0.06	0.60	0.44 – 0.71
T11	0.07	-0.06 – 0.21	0.35	-0.04 – 0.61
Mean	**0.43 ± 0.31**		**0.68 ± 0.23**	

TABLE 3: ICC values for frontal plane knee joint angles of within-day subjects and their grouped means and standard deviations.

Subject	Right Ankle ICC	CI 95%	Left Ankle ICC	CI 95%
T1	0.50	0.39 – 0.60	0.73	0.66 – 0.79
T2	0.91	0.82 – 0.95	0.71	0.62 – 0.78
T3	0.86	0.79 – 0.90	0.63	0.55 – 0.70
T4	0.19	0.01 – 0.35	0.37	0.26 – 0.47
T5	0.83	0.78 – 0.87	0.84	0.55 – 0.92
T6	0.70	0.29 – 0.85	0.52	0.28 – 0.67
T7	0.56	-0.05 – 0.80	-0.01	-0.13 – 0.10
T8	0.47	0.35 – 0.58	0.78	0.15 – 0.91
T9	0.29	0.07 – 0.47	0.02	0.08 – 0.14
T10	0.67	0.26 – 0.83	0.36	0.23 – 0.47
T11	0.79	0.66 – 0.86	0.82	0.53 – 0.91
Mean	**0.62 ± 0.22**		**0.53 ± 0.29**	

TABLE 4: ICC values for frontal plane ankle joint angles of within-day subjects and their grouped means and standard deviations.

Subject	Right Knee ICC	CI 95%	Left Knee ICC	CI 95%
T12	0.53	0.43 – 0.62	0.47	0.37 – 0.57
T13	0.81	0.15 – 0.93	0.38	0.07 – 0.59
T14	0.88	0.84 – 0.90	0.70	0.63 – 0.76
T15	0.70	0.47 – 0.81	0.13	0.00 – 0.25
T16	0.66	0.38 – 0.80	0.82	0.61 – 0.90
T17	0.72	0.65 – 0.77	0.17	0.04 – 0.29
T18	0.84	0.72 – 0.90	0.71	0.63 – 0.78
T19	0.82	0.77 – 0.85	0.18	0.05 – 0.31
T20	0.74	0.56 – 0.83	0.82	0.64 – 0.89
Mean	**0.75 ± 0.10**		**0.49 ± 0.27**	

TABLE 5: ICC values for frontal plane knee joint angles of between-day subjects and their grouped means and standard deviations.

Subject	Right Ankle ICC	CI 95%	Left Ankle ICC	CI 95%
T12	0.54	0.11 – 0.75	0.04	-0.07 – 0.17
T13	0.60	0.51 – 0.69	0.17	0.04 – 0.30
T14	0.22	0.08 – 0.34	0.74	0.61 – 0.82
T15	0.56	0.40 – 0.68	0.46	0.35 – 0.56
T16	0.50	0.41 – 0.59	0.32	0.21 – 0.42
T17	0.77	0.71 – 0.82	0.28	-0.09 – 0.60
T18	0.58	0.37 – 0.72	0.33	-0.01 – 0.57
T19	0.28	0.09 – 0.60	0.59	0.39 – 0.71
T20	0.66	0.05 – 0.85	0.08	-0.03 – 0.21
Mean	**0.53 ± 0.17**		**0.34 ± 0.26**	

TABLE 6: ICC values for frontal plane ankle joint angles of between-day subjects and their grouped means and standard deviations.

6 DISCUSSION

When comparing sagittal within-day joint angles for countermovement jump, deep learning approach yields very high ICC values for hip joint, knee joint and ankle joint. Similar values were found for between-day measurements separated by two-weeks for all measured joints. Thus, according to guidelines of Koo and Li (2016), only between-day value of ankle joint gave good test-retest reliability, and the rest demonstrated excellent test-retest reliability. Hence deep learning approach for determining sagittal hip, knee and ankle joint angles in countermovement jump seems to provide very repeatable results independent of the time between the measurements. The proposed hypothesis for sagittal joint angles can be accepted.

However, similar results were not found for the same jumps in frontal plane data. Deep learning approach gave varying ICC values for knee and ankle joint angles. Also, some of the ICC values were negative for few subjects. For within-day measurements, ICC values revealed poor (ICC = 0.43 ± 0.31) test-retest reliability for right knee angle, and moderate test-retest reliability for left knee (ICC = 0.68 ± 0.23), right ankle (ICC = 0.62 ± 0.22) and left ankle (ICC = 0.53 ± 0.29) angles. The results were similarly varied over between-day subjects, when assessing the frontal plane data of countermovement jumps. Between-day ICC values proved good (ICC = 0.75 ± 0.10) test-retest reliability for right knee angle, moderate test-retest reliability for left ankle angle (0.53 ± 0.17), and poor test-retest reliability for left knee (ICC = 0.49 ± 0.27) and right ankle (ICC = 0.34 ± 0.26) angles. It is also important to consider the high deviation of these values: Some of the subjects presented excellent test-retest reliability in some of the joint angles (as an example view left knee ICC for the subject in figure 14), and some instead presented moderate or very poor reliability values (for example right ankle of the subject in figure 14). Thus, deep learning approach may not be suitable for repeated measurements, and the proposed hypothesis for frontal plane joint angles is declined. However, the high variance in joint kinematics in frontal plane may not be due to the system, but also due to the variability of frontal plane kinematics in countermovement jump itself.

Since there were clear variations for both within-day and between-day frontal plane results, it may be concluded with high likelihood that the reason for the variation did not lie in the time period between the measurements. Instead, the reason for varying results may be due to the 1) deep learning system itself, 2) variability in countermovement jump kinematics in frontal plane, 3) poor and inadequate training of deep learning model, 4) effect of learning and fatigue on jump kinematics in between-day trials or 5) a combination of these factors. To determine if deep learning system itself is causing the variability, possible other factor (points 2-5 listed above) need to be evaluated.

In human movement there is commonly variability in joint range of motion and muscle activation and coordination strategy, depending on the task and its complexity. Indeed, when analysing 3D-kinematics of countermovement jumps there exists coordination pattern variability in countermovement jumps, and it tends to be larger for younger athletes (Raffalt et al. 2016). Wren et al. (2020) demonstrated similar large within-subject variability for drop jumps, which have quite similar movement pattern to CMJ, with variability in kinematic parameters: Their results gave variation of range of motion of measured joints between 2-11° for sagittal plane, between 2-6° for frontal plane and between 2-7° for transverse plane. Furthermore, Carroll et al. (2019) demonstrated just moderate reliability for countermovement jump depth (ICC = 0.61) for intrasession CMJ depth and poor reliability for intersession CMJ jump depth (ICC = 0.39), indicating that there would be probably exist differences also in angular values of joints. Carroll et al. (2019) also demonstrated high intra- and intersession reliability for jump height (intrasession ICC = 0.94, intersession ICC = 0.92). This study also demonstrated relatively high within- and between day reliability for jump heights (within-day ICC = 0.74, between-day ICC = 0.94). This implies that there is not high variability in jump heights and thus this factor does not have a large impact to the movement variability in CMJ.

Variability might also be affected by leg movement during the flight phase of CMJ. Firstly, small variation in the take-off might lead also to different type of muscle activation to prepare for the landing, increasing the variability of joint angles in the air between two trials. As another example, a few subjects had a large ankle dorsiflexion during the flight phase, which may have blocked the vision of ankle marker in frontal-view videos. This could increase the amount of filtering of frames where there are missing markers, and so increase the variability between two jumps.

In addition, variability might be affected by countermovement jump movement strategy: Rauch et al. (2020) demonstrated in their study how NBA players can be divided to three different eccentric phase movement strategies during CMJ (stiff flexors, hyper flexors, and hip flexors). There may theoretically be more room for variability for athletes belonging to hyper flexors and hip flexors group, since they have more angular displacement at each joint during CMJ compared to stiff flexors. The variability of joint kinematics during a countermovement jump has not been studied from frontal plane directly, but all the factors mentioned above could be explaining the variability in frontal plane kinematics to some extent in the current study.

Another factor explaining varying test-retest reliability of frontal-plane angles may be weak training of deep learning model. Discrepancies in training the model may have come from poor judgement of marker positions, compromised vision due to placement of hands over hips, different types of clothes on the subjects in the training set, and the number and variability of training images. Placement of knee, ankle and toe markers were very consistent during manual labelling thanks to their clear visibility and separability. Manual labelling of the hip joint was however a time consuming task. To stay consistent in placing the hip marker, the researchers had to look at the shape of the thigh and hip. This process was quite precise and consistent for all the subjects, who had tight sports pants without loose parts and relatively short shirt, with hand placed at the sides of the body above the hip. However, for several subjects this process was more tedious due to baggy shorts and longer shirts and/or too low placement of the hands at the side of the body. It is however important to note that training these models with as many different types of subject settings may help in creating more universal and practical model for use. This may however have caused more variability in the placement of the hip marker, resulting in larger variability for knee joint angles. This however is not the only factor affecting the end results, since great variability existed also at ankle joint angles (tables 4 and 6).

Furthermore, if the variability in frontal angles is due to poor model training, this should be evident in the tracked videos. After doing general observations on the test set on the accuracy of marker placement, this should not be the case. This fact highlights that the reason for the higher variability in frontal kinematics might be due to the variability in the countermovement jump itself and other factors, rather than poor model training. However, it is important to note that only a single model per camera view was trained. It is possible that with more data and/or different hyperparameters, the results would have been less variable.

There might be effects from learning and fatigue for between-day measurements, since they were separated by two-weeks. Since subjects of this study were actively training, they could have had a strenuous training session a few days prior to the tests, which could decrease joint range of motion and subsequently decrease the correlation between the sessions, if there were no similar stress factors before the first tests. Another factor affecting the correlations could have come from the level of the athlete and/or learning process of the test protocol.

The most likely scenario is some form of combination of these factors. However, since sagittal joint angles provided excellent test-retest reliabilities and marker placement seemed good after general observations on frontal view of the test set, it is very likely that movement variability of subjects in frontal plane angles explain largely the varying ICC values. Additionally, the varying ICC values in frontal plane angles could also be explained by the movement strategies that different planes of movement provide. In the eccentric phase of countermovement jump, the only possibility for the joint angles is to decrease their values through flexion in sagittal plane, whereas from the start of concentric phase to the take-off these joint angles are always increasing through joint extension. In frontal plane however, there is more room for movement strategy variation. For example, less experienced athlete may have knee valgus in the first trial but neutral knee position or knee varus in the second trial, if they are for example leaning their upper body more over to another side compared to their first jump. This reasoning could explain some of the negative ICC values too (i.e. substantially different frontal plane angle changes for the same subject). In future, the best solution would probably be to measure multiple countermovement jumps in a single session and determine their means to diminish the variability. Another solution to avoid problems from placing the hip marker would be to train the model so, that the markers are placed on the lateral body outline to indicate the location of joints. This would not be anatomically correct of course, but it could improve the system performance by making the point of recognition more visible by contrasting the background to the measured leg.

A few of the ICC values were negative for frontal plane joint angles. Usually, ICC values should be between 0 and 1 to define the correlation within a class of data. The ICC is calculated by using the following ratio: ICC = (variance of interest) / (total variance) = (variance of interest) / (variance of interest + unwanted variance). Because variances are calculated by performing statistical estimates, there is a possibility of poor ICC estimations, resulting in negative values. This is often due to small sample size (Liljequist et al. 2019). Negative values imply low or non-existing correlation between two sets of data.

A few limitations of this study should be acknowledged. The first source of limitations exists in the group that was studied. The results apply mostly to athletes, and thus provide information about system performance of deep learning approach for more consistent jumps with relatively small variability in joint angles. The variability of countermovement jumps might be larger for non-athletes, leading to larger inconsistencies in subsequent measurements, regardless of the measurement system used. It would be beneficial to recognise these differences of variability of human movement when assessing performance data. Additionally, there could have been more sources of error from athletes' lack of motivation for the tests, muscle soreness or fatigue from earlier training sessions leading to differences in countermovement jump kinematics and the effect of learning to perform countermovement jump in alternated movement strategy in between-day sessions.

In practical terms, these results imply that deep learning approach could provide a great tool for coaches and athletes to assess the sagittal joint angles in countermovement jumps. It would help many workers in the field to provide biomechanical analysis at an affordable price and with ease. Applications of analysing sagittal joint angles could include for example recognising the movement strategy groups and targeting training accordingly to personal needs of athletes. Analysing frontal plane joint angles with deep learning however would probably require more studies considering the naturally occurring variability in countermovement jump kinematics by studying the repeatability and validity of multiple jumps in the same session. Furthermore, to provide as accurate data as possible independent of movement strategies and variabilities within them, it would be essential to validate a deep learning approach for kinematic analysis by comparing its performance against an optoelectronic system in human movement.

In conclusion, deep learning approach provides very repeatable measurements for sagittal joint angles in countermovement jump, but not as such for frontal plane kinematics. This implies that deep learning approach provides an affordable and easy-to-access method to perform repeated measurements for 2-D motion analysis of countermovement jump and possibly other sports movements filmed from sagittal plane. However, the validation of these systems is required to further prove their accuracy and to provide reliable data for practitioners.

RESOURCES

Allard, P., A. Cappozzo, A. Lundberg & C. Vaughan. 1998. Three-Dimensional Analysis of Human Locomotion. Chichester, UK: John Wiley & Sons.

Carroll, K. M., Wagle, J. P., Sole, C. J., & Stone, M. H. 2019. Intrasession and Intersession Reliability of Countermovement Jump Testing in Division-I Volleyball Athletes. Journal of Strength & Conditioning Research (Lippincott Williams & Wilkins), 33(11), 2932–2935.

Chiari, L., Croce, U. D., Leardini, A., & Cappozzo, A. 2005. Human movement analysis using stereophotogrammetry: Part 2: Instrumental errors. Gait & Posture, 21(2), 197–211.

Colyer, S. L., Evans, M., Cosker, D. P. & Salo, A. I. T. 2018. A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System. Sports Medicine Open 4 (1), 24.

Cronin, N. J., Rantalainen, T., Ahtiainen, J. P., Hynynen, E. & Waller, B. 2019. Markerless 2D kinematic analysis of underwater running: A deep learning approach. Journal of Biomechanics 87, 75–82.

Cust, E. E., Sweeting, A. J., Ball, K., & Robertson, S. 2019. Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance. Journal of Sports Sciences, 37(5), 568–600.

Drazan, J. F., Phillips, W. T., Seethapathi, N., Hullfish, T. J., & Baxter, J. R. 2021. Moving outside the lab: Markerless motion capture accurately quantifies sagittal plane kinematics during the vertical jump. Journal of Biomechanics, 125.

Eltoukhy M, Kelly A, Kim C-Y, Jun H-P, Campbell R, Kuenze C. 2016. Validation of the Microsoft Kinect® camera system for measurement of lower extremity jump landing and squatting kinematics. Sports Biomechanics. 15(1):89-102.

Goodfellow, I., Bengio, Y., and Courville, A. 2016. Deep Learning. MIT Press.

Judson, L. J., Churchill, S. M., Barnes, A., Stone, J. A., Brookes, I. G. A., & Wheat, J. 2020. Measurement of bend sprinting kinematics with three-dimensional motion capture: a test–retest reliability study. Sports Biomechanics, 19(6), 761–777.

Kalmet, P. H. S., Sanduleanu, S., Primakov, S., Wu, G., Jochems, A., Refaee, T., Ibrahim, A., Hulst, L. v., Lambin, P., & Poeze, M. 2020. Deep learning in fracture detection: a narrative review. Acta Orthopaedica, 91(2), 215–220.

Klöpfer-Krämer, I., Brand, A., Wackerle, H., Müßig, J., Kröger, I., & Augat, P. 2020. Gait analysis - Available platforms for outcome assessment. Injury, 51, S90–S96.

Kobsar, D., Osis, S. T., Jacob, C., & Ferber, R. 2019. Validity of a novel method to measure vertical oscillation during running using a depth camera. Journal of Biomechanics, 85, 182–186.

Koo, T. K., & Li, M. Y. 2016. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of chiropractic medicine, 15(2), 155–163.

Leardini, A., Belvedere, C., Nardini, F., Sancisi, N., Conconi, M. & Parenti-Castelli, V. 2017. Kinematic models of lower limb joints for musculo-skeletal modelling and optimization in gait analysis. Journal of Biomechanics 62, 77 – 86.

Leszczewska, J., Czaprowski, D., Pawlowska, P., & Oponowicz, A. 2012. Inter-Examiner, Within-Session and Between-Session Repeatability of Kinematic Gait Parameters among Adult Subjects. Human Movement, 13(4), 337–343.

Liljequist, D., Elfving, B. & Skavberg Roaldsen, K. 2019. Intraclass correlation – A discussion and demonstration of basic features. PLoS ONE 14(7).

Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V., Mathis, M. & Bethge, M. 2018. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nat Neurosci 21, 1281–1289.

Mathis, A., Schneider, S., Lauer, J., & Mathis, M. 2020. A Primer on Motion Capture with Deep Learning: Principles, Pitfalls, and Perspectives. Neuron Volume 108, Issue 1, P44-65.

Mentiplay, B. F., Hasanki, K., Perraton, L. G., Pua, Y.-H., Charlton, P. C., & Clark, R. A. 2018. Three-dimensional assessment of squats and drop jumps using the Microsoft Xbox One Kinect: Reliability and validity. Journal of Sports Sciences, 36(19), 2202–2209.

Ota, M., Tateuchi, H., Hashiguchi, T., Kato, T., Ogino, Y., Yamagata, M., & Ichihashi, N. 2020. Verification of reliability and validity of motion analysis systems during bilateral squat using human pose tracking algorithm. Gait & Posture, 80, 62–67.

Ota, M., Tateuchi, H., Hashiguchi, T., & Ichihashi, N. 2021. Verification of validity of gait analysis systems during treadmill walking and running using human pose tracking algorithm. Gait & Posture, 85, 290–297.

Raffalt, P. C., Alkjær, T., & Simonsen, E. B. 2016. Intra- and inter-subject variation in lower limb coordination during countermovement jumps in children and adults. Human Movement Science, 46, 63–77.

Rauch J, Leidersdorf E, Reeves T, Borkan L, Elliott M, Ugrinowitsch C. 2020. Different Movement Strategies in the Countermovement Jump Amongst a Large Cohort of NBA Players. International Journal of Environmental Research and Public Health. 17(17):6394.

Robertson, D. G. E., Caldwell, G. E., Hamill, J., Kamen, G. & Saunders, N. W. 2014. Research methods in biomechanics. 2nh ed. Champaign, IL: Human Kinetics.

Schmitz, A., Mao Ye, Shapiro, R., Ruigang Yang, & Noehren, B. 2014. Accuracy and repeatability of joint angles measured using a single camera markerless motion capture system. Journal of Biomechanics, 47(2), 587–591.

Shai B.D. & Shai S. S. 2014. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press.

Sinclair, J., JohnTaylor, P., Greenhalgh, A., Edmundson, C. J., Brooks, D., & Hobbs, S. J. 2012. The Test-Retest Reliability of Anatomical Co-Ordinate Axes Definition for the Quantification of Lower Extremity Kinematics During Running. Journal of Human Kinetics, 35, 15–25.

van der Kruk, E. & Reijne, M. M. 2018. Accuracy of human motion capture systems for sport applications; state-of-the-art review. European journal of sport science 18 (6), 806 – 819.

Winter, D. A. 2009. Biomechanics and motor control of human movement. 4th ed. Hoboken, NJ: John Wiley & Sons, Inc

Wren, T. A. L., O'Callahan, B., Katzel, M. J., Zaslow, T. L., Edison, B. R., VandenBerg, C. D., Conrad-Forrest, A., & Mueske, N. M. 2020. Movement variability in pre-teen and teenage athletes performing sports related tasks. Gait & Posture, 80, 228–233.

Wright, C. J., Arnold, B. L., Coffey, T. G., & Pidcoe, P. E. 2011. Repeatability of the modified Oxford foot model during gait in healthy adults. Gait & Posture, 33(1), 108–112.