Neural Networks Estimate Muscle Force in Dynamic Conditions Better than Hill-type Muscle Models

Maria Eleni Athanasiadou[1,2], Monica A. Daley[3], and Anne D. Koelewijn[1,2]

[1]Chair of Autonomous Systems and Control, Department of Electrical Engineering, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany
[2]Machine Learning and Data Analytics Lab, Department Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany
[3]Department of Ecology and Evolutionary Biology, University of California Irvine, Irvine, CA, USA

February 5, 2024

Abstract

Hill-type muscle models are widely used, even though they do not accurately represent certain muscle mechanics. We explored neural networks to develop new muscle models. We trained neural networks to estimate muscle force from activation, muscle length, and muscle velocity. Training data was recorded using sonomicrometry, electromyography, and a tendon buckle on two muscles of guinea fowl. First, we compared the neural network to a Hill-type muscle model, using the same data for network training and model optimization. Second, we trained a network on a large dataset, in a more realistic machine learning scenario. We found that the neural networks generally yielded higher correlations and lower errors than Hill-type muscle models. Our neural network performed better when estimating forces on the muscle used for training of another bird than on a different muscle of the same bird, which could be explained by inaccuracies in activation and force scaling. We created a force-length and force-velocity relationship for the neural network and found that both amplification factors were underestimated and that both relationships were not replicated well outside of the training data distribution. We conclude that neural networks could provide an accurate alternative to Hill-type muscle models given a suitable training dataset.

1 Introduction

The human musculoskeletal system has sparked scientific interest for centuries, since knowledge of internal body forces is crucial in clinical assessments [1], prosthesis design [2], and injury prevention [3], among others. However, direct measurements of these forces require highly invasive sensors and procedures, and are therefore generally not performed in humans, especially not during movement. Instead, scientists have focused on the musculoskeletal system of other species to understand the mechanics of muscles and other tissues. For example, A. V. Hill deciphered the working of muscles using frog specimens at maximum activation [4]. Based on his research, the Hill-type muscle model has been developed, which is widely applied to answer research questions in humans (e.g., [1, 5]) and other animals, such as goats [6]. This model is a so-called phenomenological model, meaning that it mathematically describes an observed relationship, in this case between force development, muscle activation, muscle length, and muscle velocity, without necessarily describing the actual (chemical) process that happens in the muscle.

The Hill-type muscle model consists of three components representing the active muscle fibre, passive tissue in series (mainly the tendon), and passive tissues parallel to the active muscle fibres. The contractile element (CE) describes the behavior of active muscle fibre, mainly the force-length-velocity relationship [7].

This relationship is described by the sliding filament theory [8]. At a single fiber level, when a muscle is activated, crossbridges are formed between myosin and actin filaments. The formation of these crossbridges leads to the filaments sliding across one another, causing the muscle length to change and force to be generated. The efficiency of force generation is based on the overlap of the filaments and the sliding speed. Besides the CE, the Hill-type muscle models consist of a parallel elastic element (PEE), which represents the passive tissue parallel to the active muscle fibre, and a series elastic element (SEE), which represents the passive tissue connected in series to the active muscle tissue, mainly the tendon and aponeurosis [9].

Despite the widespread use of Hill-type muscle models, it is known that these models do not fully capture muscle behavior as recorded experimentally. An important aspect of muscle mechanics that the Hill-type model does not capture is the tension change in muscle following a stretch or a shortening [10]. These tension changes have been attributed to a third filament, titin [11]. Titin winds around the thin, actin filament during crossbridge formation, and thereby acts as a spring, creating force enhancement during active stretch. During active shortening, the titin filament unwinds, thereby causing force depression during shortening [11]. Other aspects that are not included in the Hill-type muscle model are changes in optimal fibre length and maximum shortening velocity with activation level. Roszek and Huijing showed that the optimal fibre length changes with the muscle activation levels [12], while Chow and Darling showed the same for the maximum shortening velocity [13]. These Hill-type muscle model limitations could be related to the fact that the model is applied to different activation levels, while it is based on experiments performed at maximum activation [4].

To overcome these limitations, in recent years, different extensions of the Hill-type muscle model (e.g. [13, 14]), as well as completely new models have been proposed (e.g., [15, 16]). For example, the history-dependent Hill-type model [14] incorporates the force enhancement or depression caused by the muscle being stretched or shortened [9, 10], respectively. Another extension includes the relationship between maximum shortening velocity and muscle stimulation [13, 17]. Instead of extending the Hill-model, Whitney et al. [18] and Millard et al. [16] have developed completely new phenomenological models to improve muscle model accuracy. These models are based on the mechanics of the three muscle filaments, actin, myosin, and titin [11, 16], whereas the Hill-type model only includes the actin and myosin filaments as part of the sliding filament theory. While both models [16, 18] outperform the Hill-type model, the force estimations could still be improved further [16], while setting the parameters of these models can be difficult [18].

Instead of further improving these existing phenomenological models based on theories of muscle mechanics, an alternative approach is to use machine learning to develop a model completely based on in-vivo muscle mechanics data gathered during different activities [19]. In theory, a large and rich enough dataset containing such data should enable us to train a neural network that replicates muscle mechanics relevant to movement and is more accurate than a model based on ex-vivo experiments. In practice, obtaining such a dataset on humans is virtually impossible, since sensors should be implanted into the muscles to record muscle length, force, and activation. However, such datasets have been collected for guinea fowl [20] and goats [6], among others. Similar to a Hill-type model, which is normalized to optimal fibre length and maximum isometric force to be applicable to different muscles and species, a neural network trained on such a dataset might also be applicable to other muscles and species, as long as it is appropriately normalized. However, to do so, it should be validated that the trained network correctly represents muscle mechanics. Previously, an artificial neural network was successfully trained to estimate muscle force from electromyography data of a cat soleus [19]. However, the relationship between electromyography and muscle force is not always direct, since the muscle length and velocity also affect force generation [21]. This relationship is especially well visible in a dataset containing perturbations, such as in the dataset by Daley and Biewener [20].

Therefore, in this paper, we developed a neural network to estimate muscle force from muscle length, velocity, and activation. We trained neural networks on existing data from guinea fowl [20]. First, we compared the accuracy of the neural network's force predictions to that of a state-of-the-art Hill-type muscle model, using the same data to train the neural network and optimize the Hill-type muscle model. Second, we investigated the accuracy of a network trained on a large dataset, in a more realistic machine learning scenario. We first evaluate the network by comparing the estimated muscle forces to the measured ones and then explored if the network replicated known muscle mechanics, specifically the force-length and force-velocity relationship. We expect that these relationships should emerge from the network if it correctly represents muscle mechanics, and could therefore provide an additional validation of the neural network's ability to fully capture muscle behavior, which is necessary to confidently use the model across species.

2 Materials and Methods

We used an existing dataset containing activation, muscle length, and muscle velocity as inputs and muscle force as output that was recorded on guinea fowl [22]. We used this dataset for network training and Hill-type muscle model optimization, and evaluated how well the network and model could estimate forces on the same muscle and bird as used for training, a different muscle of the same bird, the same muscle of a different bird, and a different muscle of a different bird. All code used for neural network training, Hill-type muscle model optimization and analysis can be found here [23].

2.1 Guinea Fowl Data

The dataset contained data from six guinea fowl. We summarize the protocol here and refer readers to Daley and Biewener [20] for a more detailed description. Briefly, this dataset contained between 12 and 15 walking and running trials of about 15 to 30 seconds for each bird. The measurements were taken during level and obstacle trials between 1.8 to 4.5 m/s. During the obstacle trials, the birds would suddenly step on a raised obstacle (5 or 7 cm) every four to five steps. Data was recorded using a tendon buckle (force), sonomicrometry (contractile element length), and electromyography (activation) for two different muscles; the lateral head of the gastrocnemius (LG) and the digital flexor to the lateral toe (DF). The contractile element velocity in the dataset was calculated as the time derivative of the contractile element length. Out of the six guinea fowl, we used the data of both muscles for three birds. For the fourth bird, the data quality of the LG was not very good due to cross-talk noise in the electromyography and the quality of the sonomicrometry for the DF was poor on day 1. Therefore, we did not include the LG data and the DF data recorded on day 1. We only used the LG data for the fifth bird, since the electromyography failed for the DF. We omitted the sixth bird entirely due to data quality issues. In previous studies (e.g. [20]), individual strides with good quality were identified for each bird, but this approach is not feasible here, because many continuous strides are required.

We used the tendon force in N, the normalized contractile element length and velocity, and the raw electromyography signal and processed the data as follows. First, we computed the maximum isometric force of each bird's LG and DF-IV using the physiological cross-sectional area [20], and assuming a maximum stress of 0.36 N/mm^2 [24]. We then normalized the measured tendon force using this maximum isometric force. Then, we processed the electromyography data similar to [18]: we first applied a third order Butterworth high-pass zero-phase digital filter with a cut-off frequency of 30 Hz, then we rectified the signal, after which we applied a third order low-pass zero-phase digital filter with a cut-off frequency of 6 Hz. This activation signal was then normalized to the maximum activation in the processed data over all trials of the specific bird and muscle, and a time delay of 23.6 ms was added to account for excitation – contraction coupling [20]. The same 6 Hz low-pass filter was also applied to the contractile element velocity.

2.2 Neural Network Training

We designed several neural networks to estimate tendon force from activation, contractile element length, and contractile element velocity. To replicate a Hill-type muscle model, these networks were set up to calculate the tendon force at each time point individually from the inputs at the same time point. First, we compared two neural networks to the Hill-type muscle model and trained them on the same data that was also used to optimize the Hill-type muscle model parameters. Due to the computational cost related to the model optimization, we made this comparison on a smaller dataset, consisting of a single trial. Specifically, we trained a network using the data of the LG of the first trial of bird 1, a level trial at 1.8 m/s. The second network was trained on the twelfth trial of the same bird and muscle, a trial with 7 cm obstacles at 4.5 m/s. Second, we trained a neural network on a large dataset to investigate its prediction accuracy in a more commonly used machine learning scenario. To do so, we trained this network on LG data of 43 trials of birds 1-3.

We used the following process to train the networks. First, we split the data into three categories; the *training* data, on which the neural network was trained, the *validation* data, to validate the hyperparameter optimization, and the *test* data, on which the trained neural network made predictions. The training data contained the first 80% of the trials, while the validation was performed on the last 20%. This split was

chosen to ensure that the training data did not include information from a later time point than any of the validation data, which could lead to data leakage [25]. We normalized the complete training dataset, to have a mean of zero and a standard deviation of one. We optimized the network's hyperparameters, specifically the number of layers, number of nodes in each layer, and the activation function of the hidden layers, using the automatic hyperparameter optimization that is included in the MATLAB function `fitrnet`, which uses a Bayesian optimization. We repeated this hyperparameter optimization 10 times when training on a single trial, and five times when training on the large dataset, to account for random effects of the Bayesian optimization, and chose the network with the smallest loss overall. We performed all training and analysis in MATLAB R2022a and R2022b (Mathworks, Natick, MA, USA).

2.3 Hill-type Muscle Model and Optimization

We used the Hill-type muscle model described by Zajac [7], Van den Bogert et al. [26], and Herzog [27]. We chose this model, since its mathematical equations are such that the shape of the force-length and force-velocity relationship are preserved during the optimization. We found that for other models, the equations could be optimized such that the shape of the force-length and force-velocity relationship changed. We determined the tendon force, F_{Hill}, as the sum of the force in the contractile element, F_{CE}, and the parallel elastic element, F_{PEE}, using the following relationships:

$$F_{Hill} = F_{CE} + F_{PEE} \tag{1}$$

$$\text{with} \quad F_{CE} = af(l_{CE})g(v_{CE})F_{max} \tag{2}$$

$$f(l_{CE}) = \exp\left(-\left(\frac{l_{CE}-1}{W}\right)^2\right), \tag{3}$$

$$g(v_{CE}) = \begin{cases} \frac{v_{CE,max}+v_{CE}}{v_{CE,max}-v_{CE}/A_{rel}} & \text{if } v_{CE} \leq 0, \\ \frac{g_{max}v_{CE}+c_3}{v_{CE}+c_3} & \text{if } v_{CE} > 0 \end{cases} \tag{4}$$

$$c_3 = v_{CE,max}A_{rel}\frac{g_{max}-1}{A_{rel}+1}, \tag{5}$$

$$F_{PEE} = k_{PEE}(l_{CE} - l_{slack,PEE})^2, \tag{6}$$

where $f(l_{CE})$ is the force-length relationship with width, W, $g(v_{CE})$ the force-velocity relationship with maximum shortening velocity, $v_{CE,max}$, the curvature of the force-velocity relationship, A_{rel}, the maximum force amplification during lengthening, g_{max}, and factor c_3, which was calculated such that the force-velocity relationship is continuous, k_{PEE} the stiffness of the parallel elastic element, and $l_{slack,PEE}$ its slack length. Note that the muscle length, l_{CE} and velocity, v_{CE}, are normalized to optimal fibre length,

We optimized the Hill-type muscle model using the same data as used for network training to allow for a fair comparison between the model and the neural network. We optimized the width of the force-length relationship, W, the maximum shortening velocity, $v_{CE,max}$, the curvature of the force-velocity relationship, A_{rel}, the maximum force amplification during lengthening, g_{max}, the stiffness of the parallel elastic element, k_{PEE}, and its slack length, $l_{slack,PEE}$. As an objective function, we used the mean square difference between the calculated, F_{Hill}, and measured, F_{meas}, muscle force:

$$J = \frac{1}{N}\sum_{i=1}^{N}(F_{meas}(i) - F_{Hill}(i))^2 \tag{7}$$

We optimized the parameters using covariance matrix adaptation, evolutionary strategy (CMA-ES) [28], using 50 iterations. We performed the optimizations in MATLAB R2021a (Mathworks, Natick, MA, USA).

2.4 Data Analysis

First, we investigated if the neural network was able to outperform the Hill-type muscle model. To do so, we optimized the Hill-type muscle model on the same data of bird 1 as was used for neural network training. We trained two neural networks and optimized the Hill-type muscle model twice: once using the first trial at 1.8

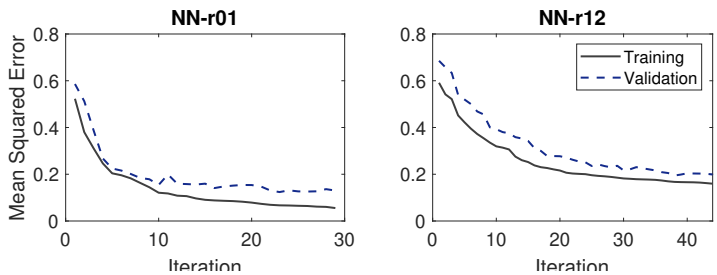

Figure 1: Training and validation loss during training of the neural networks NN-r01 (level at 1.8 m/s) and NN-r12 (7 cm incline at 4.5 m/s).

m/s and a level surface and once using the twelfth trial at 4.5 m/s with 7 cm obstacles. We call these models NN-r01, NN-r12, Hill-r01, and Hill-r12, respectively. We compared all models using the root-mean-square error (RMSE) and the coefficient of determination (R^2) between the measured and estimated (either through the neural network or Hill-type muscle model) force. We compared both metrics on the same bird and muscle as used for training, on all trails that were not used in training, on the same bird's DF, to investigate both models' ability to predict force of another muscle, on bird 2's LG, to investigate the models' ability to predict force of another bird, and on bird 2's DF, to investigate the prediction accuracy when both bird and muscle were different.

Second, we investigated the neural network accuracy in a more common training approach with a large dataset. We used the recordings of the LG of all trials of birds 1-3 to train the neural network, such that we could test on high quality data of an unseen bird for the LG (bird 5) and the DF (bird 4). Then, we used a similar approach as before and evaluated the network's estimation accuracy on a bird and muscle as used in training, on the same bird but another muscle (DF), on the same muscle of a different bird (bird 5), and on another muscle (DF) of a different bird (bird 4).

Third, to investigate if the final network replicated muscle mechanics, we compared the force-length and force-velocity relationship of the neural network trained on a large amount of data to those of the Hill-type muscle model. To do so, we varied the fiber length between 60% and 140% of the optimal fibre length and calculated the resulting force normalized to maximum isometric force, assuming isometric conditions. The force-length relationship includes the PEE, since the CE and the PEE cannot be separated in the neural network. Then, we varied the contractile element velocity between -10 and 10 fibre lengths per seconds, assuming optimal fibre length, and again calculated the resulting force normalized to maximum isometric force. We repeated these for five different activation levels: 0.2, 0.4, 0.6, 0.8 and 1.

3 Results

3.1 Comparison of Neural Network to Hill-type Muscle Model

We found that the training and validation loss for NN-r01 were slightly lower than for NN-r12, while for both models, the training and validation loss show a similar trend (Fig. 1). After training, the neural network NN-r01 had a training loss of 0.055 and a validation loss of 0.13 after 29 iterations, while NN-r12 had a training loss of 0.16 and a validation loss of 0.20 after 44 iterations. Both models consisted of three hidden layers with rectified linear activation units, NN-r01 with 300, 254, and 200 layers, and NN-r12 with 188, 140, and 13 nodes. We also found that similar results were achieved with vastly different network architectures, as the best networks found during hyperparameter optimization had similar losses but different characteristic features. For example, for NN-r12, the number of layers varied between two and three, the first layer ranged from 21 to 291 nodes, the second layer ranged between 10 and 297 nodes and the third layer between 13 and 245 nodes. Figure 1 also shows some evidence of overfitting to the training data, as the training loss is lower than the validation loss.

Our neural networks yielded lower RMSEs than the Hill-type muscle models for all four test cases: when applied to the bird and muscle that were used for training and optimization, when applied to a different muscle, to a different bird, and when both are different (Table 1). Both neural networks produced lower

Table 1: Root mean square erros (RMSEs), normalized to maximum isometric force, between measured and estimated forces, averaged over all trials in the testing data. Bird 1 LG represents the same bird and muscle (lateral gastrocnemius, LG) as used for training, bird 1 DF the same bird, but a different muscle (digital flexor, DF), bird 2 LG a different bird, but same muscle, and bird 2 DF a different bird and a different muscle. Bold font indicates lowest RMSE.

	NN-r01	NN-r12	Hill-r01	Hill-r12
Bird 1 LG	0.0106	**0.0102**	0.0174	0.0179
Bird 1 DF	0.0690	**0.0603**	0.1049	0.1209
Bird 2 LG	0.0288	**0.0259**	0.0375	0.0329
Bird 2 DF	0.0793	**0.0709**	0.1058	0.1455

Table 2: Coefficient of determination (R^2) between measured and estimated forces of the LG measurements for the same four test cases: the same bird and muscle as for training, the same bird but a different muscle than used for training, a different bird but the same muscle as used for training, and a different bird and different muscle. Bold font indicates highest coefficient of determination.

	NN-r01	NN-r12	Hill-r01	Hill-r12
Bird 1 LG	0.76	**0.76**	0.34	0.34
Bird 1 DF	0.36	**0.52**	0.16	0.024
Bird 2 LG	0.61	**0.66**	0.35	0.46
Bird 2 DF	0.17	**0.34**	0	0

RMSE values than the Hill-type muscle models, with the error reduction between around 20%-30% for the LG and 30%-50% for the DF. The neural network NN-r12 had the lowest RMSE between the two neural networks on all four test cases. For the Hill-type muscle models, the RMSE of Hill-r01 was lower, except when the models were tested on the same muscle of a different bird (bird 2's LG).

Similar to the superior performance in RMSE, the coefficient of determination of our neural networks was also higher than those of the Hill-type muscle models (Table 2). In all test cases, the neural network NN-r12 performed best, while the Hill-type muscle model Hill-r01 performed worst, except for the test on the different muscle of the same bird, where the model Hill-r12 performed worst. The coefficient of determination was zero for both Hill-type muscle models for the test case with a different muscle of a different bird.

We found that the neural network estimation matched the measurements for all test cases, but especially well on the same bird and same muscle, indicating that there could be some overfitting to the training data (Fig. 2). Especially in Fig. 2(a) the neural network's force estimates matched the measured data very well. The peak force in the DF muscle is underestimated by the neural network, while the Hill-type muscle model overestimated the force in this muscle (Fig. 2(b) and (d)). At the same time, the duration of the individual force peaks is longer in the neural network estimations than in the measurements. Using the LG muscle of a different bird, the Hill-type muscle model estimations were less smooth than the measurements and the neural network estimations, but the peaks' magnitudes matched better (Fig. 2(c)).

3.2 Predictions with Neural Network Trained on Large Dataset

We found that the neural network trained and tested on a large dataset combining different birds had a training loss of 0.13 and a validation loss of 0.15 after 106 iterations (Fig. 3). This validation loss was similar to the loss of NN-r01, while the training loss was slightly higher. The network trained on the large dataset also consisted of three layers with rectified linear activation units. The first layer had 293 nodes, the second 164 and the third 52. Again, similar results could be achieved with different network architectures, with between one and three layers, and between one to 300 nodes in each layer. However, the difference between the training and validation loss was smaller with the larger dataset compared to a smaller dataset, indicating less overfitting to the training data, while the difference between training and validation loss again remained

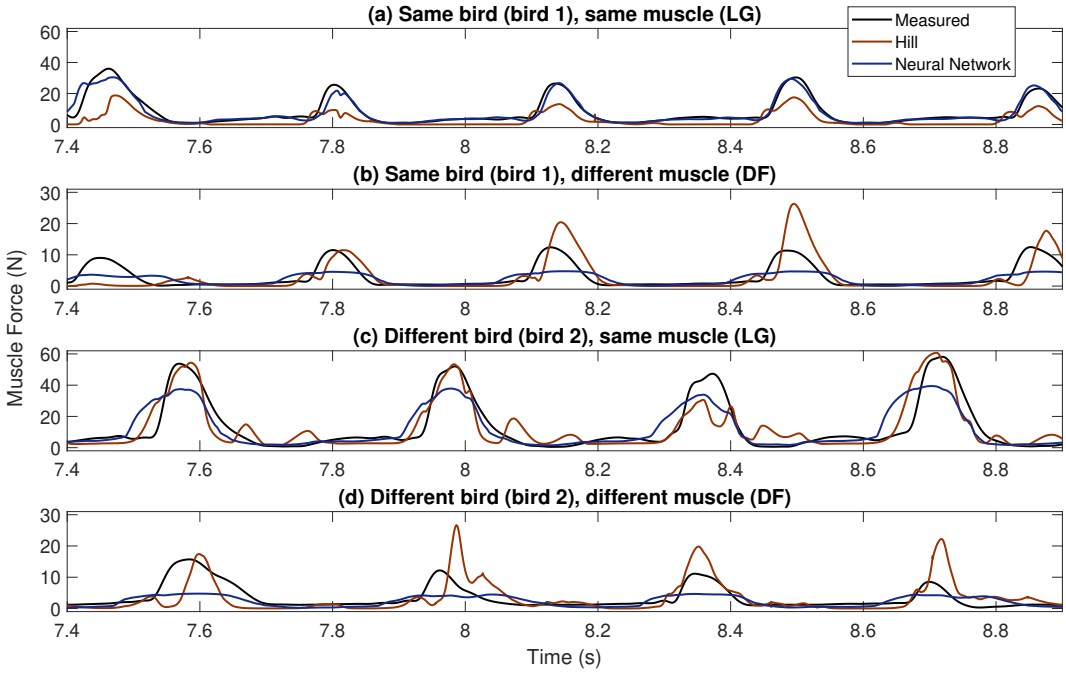

Figure 2: Tendon force over time for part of a trial with speed 3.8 m s^{-1} and 7 cm elevation for (a) the same bird and muscle as used for network training and Hill=type muscle model optimization, (b) a different muscle of the same bird, (c) the same muscle of a different bird, and (d) for a different muscle of a different bird. We plotted the results for the neural networks and Hill-type muscle models with the lowest RMSE, i.e., NN-r12 in (a)-(d), Hill-r01 in (a), (b), and (d), and Hill-r12 in (c).

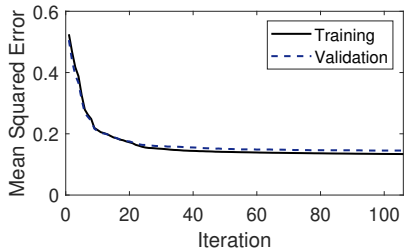

Figure 3: Training and validation loss during training of the neural network for the large dataset.

similar during the training.

Our network predicted muscle force more accurately when predicting on the same muscle of a different bird compared to predicting on a different muscle of the same or a different bird, since the RMSE is higher and the coefficient of determination is lower for the DF than for the LG (Table 3). Unsurprisingly, the lowest RMSE and highest coefficient of determination were achieved when using the training data (birds 1-3 LG). For force estimations on the same muscle, we found an increase in RMSE of at least 0.004 and a decrease in coefficient of determination of at least 0.62 for bird 5. For force estimations on a different muscle of the same bird, we found an increase in RMSE of at least 0.034 and a decrease in coefficient of determination of at least 0.2 for the best performing test bird (DF, Bird 1). We also found that the RMSE for the other muscle (DF) was similar between the birds used for training and the one not used for training, but at least three times higher than when the same muscle (LG) was used. However, the coefficient of determination was generally larger for the unseen muscle (DF) of the same birds (birds 1-3) than for LG muscle of the unseen bird (bird 5).

Table 3: RMSEs, normalized to maximum isometric force, and coefficient of determination (R^2) between force estimations and measurements for the neural network trained on the LG data of birds 1-3.

	LG				DF			
	Bird 1	Bird 2	Bird 3	Bird 5	Bird 1	Bird 2	Bird 3	Bird 4
RMSE	0.0096	0.020	0.010	0.024	0.054	0.065	0.19	0.13
R^2	0.80	0.80	0.92	0.18	0.60	0.45	0.048	0.098

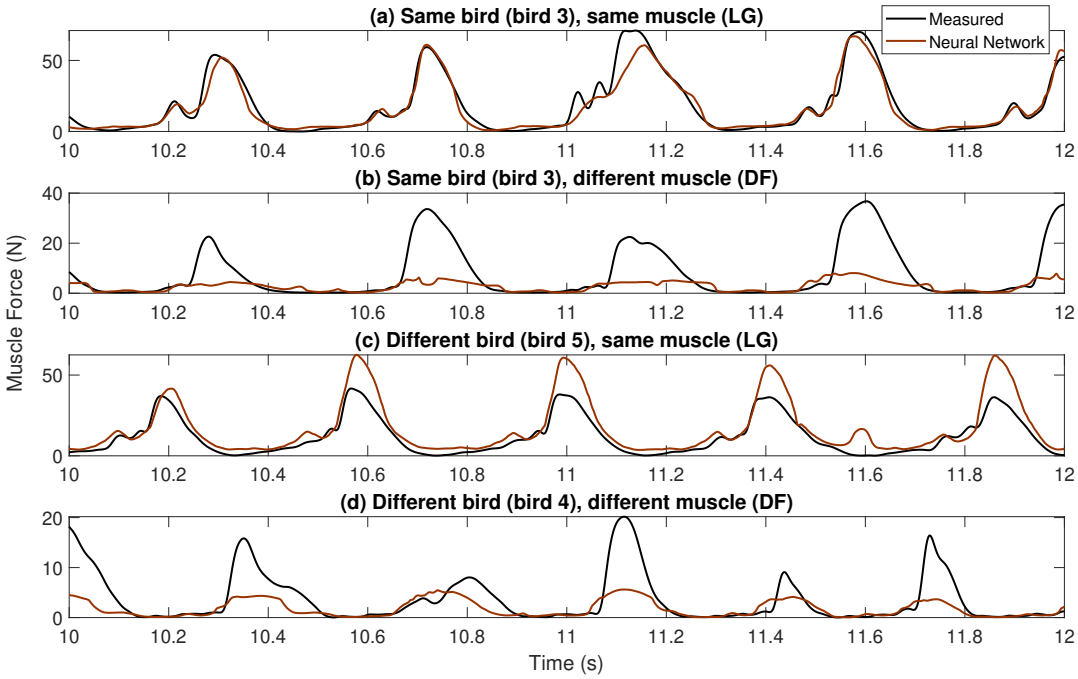

Figure 4: Tendon force over time for part of a trial with speed 3.5 m s^{-1} and 7 cm elevation for (a) the same bird and muscle as used for network training, (b) the same bird but a different muscle, (c) the same muscle of a different bird, and (d) a different muscle of a different bird.

Our neural network trained on a large dataset was again able to replicate the measured force well when applied to a muscle it was trained on, while for the other muscle, the peak force was underestimated (Fig. 4). Figure 4(a) shows data that was used for training, and here the neural network estimations matched the measurements very closely, while for a different bird, the peak forces were overestimated and there was less similarity in the patterns (Fig. 4(c)). For example at 10.1, 11.3, and 11.75 seconds in Fig. 4(c), the smaller peaks are similar, while the larger peaks (e.g., 10.6, 11.0, 11.4 seconds) are overestimated. For the DF, which was not used in training, the neural network underestimated the peak force by around 15 N.

A visualization of our neural network's force-length and force-velocity relationships can explain the underestimation of the DF force, and shows that the sliding filament theory was not represented well by the network (Fig. 5). The maximum normalized muscle force was about 0.3 for both relationships, which indicates that the maximum muscle force is never achieved. This result indicates that the normalization to maximum activation in the processed data might not have been equivalent between the DF and LG, which explains the underestimations seen in Fig. 2 and Fig. 4. At 0.2 activation, the force-length relationship has a characteristic shape with a maximum at close to optimal fibre length, but with increasing activation levels, the maximum shifts to increasingly longer fibre lengths. At maximum activation, the descending slope is not visible anymore. The force-velocity relationship shows a decrease in muscle force generation with increasing

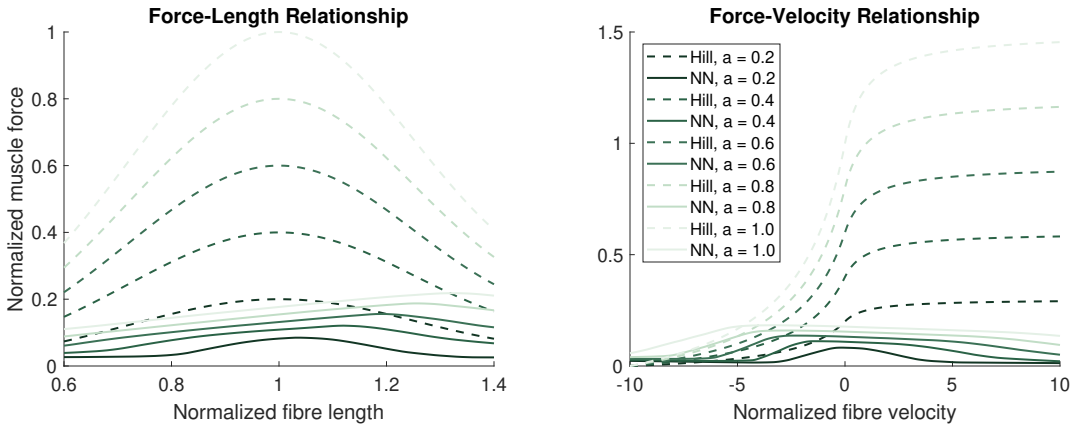

Figure 5: Force-length relationship, including parallel elastic element, and force-velocity relationship for the neural network and the Hill-type muscle model.

shortening velocities, but the increased force generation during lengthening is not visible.

4 Discussion

We trained neural networks to estimate muscle force from activation, CE length, and CE velocity and investigated if these networks can estimate muscle forces more accurately than a Hill-type model, and how well they perform on unseen birds and muscles. We found that the neural networks generally outperformed Hill-type muscle models that were optimized using the same data as used for network training. The RMSEs were lower for all test cases (Table 1), while coefficients of determination were higher (Table 2). In general, the validation and testing errors were of similar magnitude independent of dataset size used for training, but the smaller dataset led to more overfitting, as the difference between training and validation error was higher. We also found, both for the networks trained on smaller and larger datasets, that the neural network generally estimated force more accurately (lower RMSE and higher coefficient of determination) when applied to the same muscle as used for training of a different bird, compared to estimations on a different muscle of a bird used for training (Fig. 2, Fig. 4). It seemed that performance for another muscle was worse due to inaccuracies in the maximum isometric force or in the activation scaling, which led to force underestimations when applied on another muscle (Fig. 2, Fig. 4). This underestimation could also explain why the maximum force amplification in the force-length and force-velocity relationships was only 0.3, instead of 1 for the force-length relationship, and about 1.5 for the force-velocity relationship (Fig. 5). Furthermore, we found that our network only represented the force-length relationship at low activation levels, while it represented the force-velocity relationship well only for shortening but not lengthening, which indicates that the network did not perform well outside of the training data distribution (Fig. 5). A dataset that better represents the complete functional range could improve network performance, while a physics-informed neural network (e.g. [29, 30]) could also improve network performance at extremes where data recording is difficult. A physics-informed neural network calculates the loss not only from training data, but also based on some physics relationships that the network should follow. For example, in case of a muscle model, a loss could be included based on how well the network follows the force-length and force-velocity relationship.

Our results show the potential of using neural networks as a tool to estimate muscle forces from muscle activation, muscle length, and muscle velocity. By training the neural networks on datasets that were recorded in dynamic conditions, we can ensure that the network better captures dynamic muscle behaviour than Hill-type muscle models, which are based on limited, static conditions. These improved estimations are highlighted in Fig. 4, where the neural network estimates the smaller peaks that occur just before large peaks (e.g., at 10.6, 11.0, 11.4 s in (a)), which are generally not captured by Hill-type muscle models. These behaviours will be especially important in faster movements, such as running or change-of-direction movements, where the dynamic behaviour of muscles is important [31]. However, to be able to implement a neural network, trained on guinea fowl data, to estimate muscle forces in a human musculoskeletal model,

we need to have confidence that the network represents muscle mechanics correctly.

However, our network did not represent the sliding filament theory well, since the force-length and force-velocity relationship were not represented well (Fig. 5). First, the maximum amplification factors in the force-length and force-velocity relationships (around 0.3, see Fig. 5) were much lower than expected, which could indicate that the activation scaling, to the maximum of the processed data, was not equivalent between the LG and the DF or that the maximum isometric force was not accurately estimated. Second, we found that the change of the force-length relationship with activation was not correct, while for the force-velocity relationship, it was correct. For our network, the optimal fibre length increased with increasing activation, while previous experiments have shown that it decreases with increasing activation (e.g., [12, 32]). However, the maximum shortening velocity decreased with decreasing activation, as was previously shown [13]. Third, the network's force-length-velocity relationship did not display the correct shape for the entire tested range. Only at low activation was the optimal fibre length displayed accurately, around 1 (optimal fibre length), while for larger activation, it was longer. Similarly, the force-velocity relationship did not display a force increase at lengthening velocities, though the shape for shortening velocities was relatively accurate. The network's output was inaccurate at the extremes of the training data, e.g., at large activation and large lengthening velocities. Therefore, this inaccuracy could be explained by a lack of training data in these regions, since most data was recorded at low and medium activation, and fibre lengths and velocities closer to optimal and isometric conditions. Another reason could be found in a recent study by Schwaner et al. [32], who found that in-vivo force-length dynamics might not be represented well by in-situ measurements of the force-length relationship. Therefore, we need to find additional approaches to further validate the neural networks for application across species.

To achieve better neural network performance, we recommend creating a more suitable training dataset. An ideal dataset, that is applicable to a broad range of activities and species, would include data of at least three different muscles and at least three different animal species, and includes different tasks that represent the full functional range of the animals as much as possible. By including three muscles, we can perform training on two, and test on the third muscle, which would avoid a bias in the network towards the single muscle used in training. Such a bias is likely present in our network, which performed better on the muscle that it was trained on (LG) than the unseen muscle (DF). Furthermore, to apply such a network to a human musculoskeletal model, we need to at least validate the model on different species to test how well the model performs on an unseen species. It would be best to directly validate on human data, but such a validation is impossible due to the need for invasive sensors. Again, by including at least three species, we can train on two and test on the third, and thereby avoid that the network biases towards a specific species. This ideal dataset should also include the full functional range of the animals as much as possible. To do so, both submaximal and maximal tasks should be included. These maximal tasks should lead to all measured muscles being maximally activated, such that the normalization is comparable between muscles. Furthermore, it should include recordings with a large distribution of muscle lengths, and shortening and lengthening velocities. This ideal dataset, to train a network as a replacement of the Hill-type muscle model, should not include data of fatigued muscles, since fatigue alters the relationship between activation and force over time. Eventually, fatigue could also be included in a neural network if a fatigue measurement can be included in the training data.

It is also important to realize that a neural network is only as good as the data that is used for training. In practice, training datasets will always have limitations. For example, it could be that a neural network is well-suited to predict forces for walking when it is trained only on submaximal activation data, but not for jumping. Furthermore, there might be too much difference in data recordings between species and muscles to train a neural network that is applicable on all species and muscles from one large dataset. Therefore, a neural network might not be as flexible as a phenomenological model such as the Hill-type muscle model, though Hill-type muscle models are actually often applied even though they are known to be inaccurate (e.g., [10]). However, the internal relationships of a Hill-type muscle model are understood, and it is therefore known that they still represent muscle mechanics to some extent. In contrast, we do generally not understand the internal relationships of a neural network, which acts like a black box, and therefore a neural network cannot be deployed without proper validation. At the same time, DeepLabCut, an effective machine learning model for pose estimation, has a similar black box behaviour. This model can be used for different movements as well as different species through transfer learning with little additional training data [33]. Such an approach could potentially also be used to transfer muscle models between species. However, transfer learning towards

humans would again be challenging since direct measurements are impossible. Instead, it might be possible to create sufficient training data by estimating muscle forces through inverse kinematics and dynamics from optical motion capture, possibly combined with surface electromyography. In this approach, a Hill-type muscle model is normally used to estimate the muscle forces, which means that the neural network might in the end be trained to replicate a Hill-type muscle model instead of the actual muscle mechanics.

We also compared our results to other relevant muscle models [18, 19]. Liu et al. [19] successfully trained an artificial neural network to estimate muscle force in a cat soleus solely from electromyography signals [19] with lower errors than ours. When we used our data to train a model to estimate force from electromyography data only, the network performance was worse. This reduction could be caused by the fact that our dataset, which included perturbations, was specifically designed to include cases where a different force was created despite having the same muscle activation, while Liu et al. [19] used steady-state walking data, where the relationship between muscle activation and force is more direct. Due to this training approach, their model will likely not generalize to non-steady conditions, as these are not included in their training data. We also compared our coefficient of determination values to a recently developed winding filament model [18], which was optimized on the same data. Compared to this model, our coefficients of determination were comparable or higher for the LG data used in training (0.80, 0.80, and 0.92 compared to 0.81, 0.71, and 0.78 in Table 2 in [18]), while it was much lower for the unseen bird (0.18 compared to 0.74 in Table 2 in [18]). Also for the DF, the unseen muscle, our coefficients of determination were worse (0.60, 0.45, 0.048, and 0.098 compared to 0.82, 0.64, 0.64, and 0.48 in [18]). However, Whitney et al. [18] performed a model parameter optimization for each muscle and each trial individually, while we tested on unseen birds or muscles.

Since the dataset used here was not recorded specifically for neural network training, we could not test different approaches. For example, we suspect that the under- and overestimation of peak forces in the DF is related to a difference in activation scaling between the two muscles or inaccurate estimates of the maximum isometric force. We scaled the activation to the maximum activation recorded in the respective muscle over all trials of the same bird, but it is well possible that this is not equivalent between the muscles if less activation is required in one muscle than in the other for the recorded activities (e.g., 40% of the maximum for the LG and 80% of the maximum for the DF). Furthermore, the maximum isometric force was determined from the physiological cross-sectional area through the maximum stress, but the maximum stress might be different between muscles [34]. Since the network is only trained on one muscle, it has likely developed a bias towards this muscle's error in the activation and maximum isometric force, and therefore the RMSE is much lower for the LG than for the DF. Furthermore, this bias seems persistent among the same muscle of different birds, but not between different muscles of the same bird. When training on two or more muscles, such a bias could be avoided. However, since data of only two muscles was available, we could not have tested this scenario on independent data.

In this work, we trained a neural network to be similar to a Hill-type muscle model, but in the future, different network architectures could be explored to further improve network performance and have it better represent muscle mechanics. For example, there is a history dependence in muscle force generation [35], which could be captured in a recurrent neural network. Furthermore, neural networks could also be used as a tool for data-based muscle mechanics exploration. By comparing the quality of different architectures or input combinations, the relevance of different muscle states or parameters to muscle force generation could be explored to formulate hypotheses about muscle mechanics. These hypotheses could then be further tested in specific experiments.

In conclusion, we found that a neural network can estimate muscle force more accurately than a Hill-type muscle model, both on different muscles and birds, and independent of training dataset size. However, a larger dataset seemed to reduce overfitting towards the training data. We also found that, when training on data of one muscle, the network makes better estimations on the same muscle of another bird, than on another muscle of the same or a different bird. Future studies should further investigate if this effect can be mitigated by training on more than one muscle, which reduces the bias towards the muscle used for training, while we have also made recommendations on an ideal dataset including maximum activation tasks which could also mitigate this issue. In conclusion, we have shown that neural networks are a promising approach to developing new muscle models, and they could also be a helpful tool when creating new hypotheses about muscle mechanics.

References

[1] KM Steele, DL Damiano, MN Eek, M Unger, and SL Delp. Characteristics associated with improved knee extension after strength training for individuals with cerebral palsy and crouch gait. *Journal of pediatric rehabilitation medicine*, 5(2):99–106, 2012.

[2] Abbie E Ferris, Jennifer M Aldridge, Christopher A Rábago, and Jason M Wilken. Evaluation of a powered ankle-foot prosthetic system during walking. *Archives of physical medicine and rehabilitation*, 93(11):1911–1918, 2012.

[3] Emily S Matijevich, Lauren M Branscombe, Leon R Scott, and Karl E Zelik. Ground reaction force metrics are not strongly correlated with tibial bone load when running across speeds and slopes: Implications for science, sport and wearable tech. *PloS one*, 14(1):e0210000, 2019.

[4] Archibald Vivian Hill. The heat of shortening and the dynamic constants of muscle. *Proceedings of the Royal Society of London. Series B-Biological Sciences*, 126(843):136–195, 1938.

[5] Anne D Koelewijn and Jessica C Selinger. Predictive simulations to replicate human gait adaptations and energetics with exoskeletons. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:1931–1940, 2022.

[6] Sabrina SM Lee, Allison S Arnold, Maria de Boef Miara, Andrew A Biewener, and James M Wakeling. Accuracy of gastrocnemius muscles forces in walking and running goats predicted by one-element and two-element hill-type models. *Journal of biomechanics*, 46(13):2288–2295, 2013.

[7] Felix E Zajac. Muscle and tendon: properties, models, scaling, and application to biomechanics and motor control. *Critical reviews in biomedical engineering*, 17(4):359–411, 1989.

[8] AM Gordon, Andrew F Huxley, and FJ Julian. The variation in isometric tension with sarcomere length in vertebrate muscle fibres. *The Journal of physiology*, 184(1):170–192, 1966.

[9] Ross H. Miller. *Hill-Based Muscle Modeling*, pages 1–22. Springer International Publishing, Cham, 2018.

[10] BC Abbott and XM Aubert. The force exerted by active striated muscle during and after change of length. *The Journal of physiology*, 117(1):77, 1952.

[11] Kiisa C Nishikawa, Jenna A Monroy, Theodore E Uyeno, Sang Hoon Yeo, Dinesh K Pai, and Stan L Lindstedt. Is titin a 'winding filament'? a new twist on muscle contraction. *Proceedings of the royal society B: Biological sciences*, 279(1730):981–990, 2012.

[12] Boris Roszek and Peter A Huijing. Stimulation frequency history alters length-force characteristics of fully recruited rat muscle. *Journal of electromyography and kinesiology*, 7(3):161–177, 1997.

[13] John W Chow and Warren G Darling. The maximum shortening velocity of muscle should be scaled with activation. *Journal of Applied Physiology*, 86(3):1025–1031, 1999.

[14] Craig P McGowan, RR Neptune, and W Herzog. A phenomenological muscle model to assess history dependent effects in human movement. *Journal of biomechanics*, 46(1):151–157, 2013.

[15] Uzma Tahir, Anthony L Hessel, Eric R Lockwood, John T Tester, Zhixiu Han, Daniel J Rivera, Kaitlyn L Covey, Thomas G Huck, Nicole A Rice, and Kiisa C Nishikawa. Case study: a bio-inspired control algorithm for a robotic foot-ankle prosthesis provides adaptive control of level walking and stair ascent. *Frontiers in Robotics and AI*, 5:36, 2018.

[16] Matthew Millard, David W Franklin, and Walter Herzog. A three filament mechanistic model of musculotendon force and impedance. *bioRxiv*, pages 2023–03, 2023.

[17] Marlies Nitschke, Eva Dorschky, Dieter Heinrich, Heiko Schlarb, Bjoern M Eskofier, Anne D Koelewijn, and Antonie J Van den Bogert. Efficient trajectory optimization for curved running using a 3d musculoskeletal model with implicit dynamics. *Scientific reports*, 10(1):1–12, 2020.

[18] Christopher Whitney, Siwoo Jeong, Monica Daley, and Kiisa Nishikawa. Predicting in vivo forces of guinea fowl hindlimb muscles during perturbed locomotion using a titin-inspired model. *Available at SSRN 4153409*, 2022.

[19] Ming Ming Liu, Walter Herzog, and Hans HCM Savelberg. Dynamic muscle force predictions from emg: an artificial neural network approach. *Journal of electromyography and kinesiology*, 9(6):391–400, 1999.

[20] Monica A Daley and Andrew A Biewener. Leg muscles that mediate stability: mechanics and control of two distal extensor muscles during obstacle negotiation in the guinea fowl. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1570):1580–1591, 2011.

[21] Kiisa Nishikawa, Andrew A Biewener, Peter Aerts, Anna N Ahn, Hillel J Chiel, Monica A Daley, Thomas L Daniel, Robert J Full, Melina E Hale, Tyson L Hedrick, et al. Neuromechanics: an integrative approach for understanding motor control. *Integrative and comparative biology*, 47(1):16–54, 2007.

[22] Maria Eleni Athanasiadou, Monica Daley, and Anne D Koelewijn. Data from: Neural networks estimate muscle force in dynamic conditions better than hill-type muscle models [dataset], 2024. URL `https://doi.org/10.5061/dryad.547d7wmg8`. Dryad.

[23] Maria Eleni Athanasiadou, Monica Daley, and Anne D Koelewijn. Code used for: Neural networks estimate muscle force in dynamic conditions better than hill-type muscle models., 2024. URL `https://doi.org/10.5281/zenodo.10618728`. Zenodo.

[24] SM Cox, KL Easton, M Cromie Lear, RL Marsh, SL Delp, and J Rubenson. The interaction of compliance and activation on the force-length operating range and force generating capacity of skeletal muscle: a computational study using a guinea fowl musculoskeletal model. *Integrative Organismal Biology*, 1(1):obz022, 2019.

[25] Eva Dorschky, Valentina Camomilla, Jesse Davis, Peter Federolf, Jasper Reenalda, and Anne D Koelewijn. Perspective on "in the wild" movement analysis using machine learning. *Human movement science*, 87:103042, 2023.

[26] Antonie J Van den Bogert, Dimitra Blana, and Dieter Heinrich. Implicit methods for efficient musculoskeletal simulation and optimal control. *Procedia Iutam*, 2:297–316, 2011.

[27] Walter Herzog. Muscle. In Benno M Nigg, editor, *Biomechanics of the Musculoskeletal System*, pages 148–188. Wiley, New York, New York, USA, 2 edition, 1999.

[28] Nikolaus Hansen and Andreas Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation*, 9(2):159–195, 2001.

[29] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. 378:686–707.

[30] Jie Zhang, Yihui Zhao, Fergus Shone, Zhenhong Li, Alejandro F Frangi, Sheng Quan Xie, and Zhi-Qiang Zhang. Physics-informed deep learning for musculoskeletal modeling: Predicting muscle forces and joint kinematics from surface emg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:484–493, 2022.

[31] Friedl De Groote, Allison L Kinney, Anil V Rao, and Benjamin J Fregly. Evaluation of direct collocation optimal control problem formulations for solving the muscle redundancy problem. *Annals of biomedical engineering*, 44:2922–2936, 2016.

[32] M Janneke Schwaner, Dean L Mayfield, Manny Azizi, and Monica A Daley. Linking in vivo muscle dynamics to in situ force-length and force-velocity reveals that guinea fowl lateral gastrocnemius operates at shorter than optimal lengths. *bioRxiv*, pages 2023–10, 2023.

[33] Tanmay Nath, Alexander Mathis, An Chi Chen, Amir Patel, Matthias Bethge, and Mackenzie Weygandt Mathis. Using deeplabcut for 3d markerless pose estimation across species and behaviors. *Nature protocols*, 14(7):2152–2176, 2019.

[34] Constantinos N Maganaris, Vasilios Baltzopoulos, Derek Ball, and Anthony J Sargeant. In vivo specific tension of human skeletal muscle. *Journal of applied physiology*, 90(3):865–872, 2001.

[35] Walter Herzog. History dependence of skeletal muscle force production: implications for movement control. *Human movement science*, 23(5):591–604, 2004.