Utilizing Neural Networks and Wearables to Quantify Hip Joint Angles and Moments During Walking and Stair Ascent

Megan V. McCabe
Megan.V.McCabe.20@dartmouth.edu

Follow this and additional works at: https://digitalcommons.dartmouth.edu/engs88

Part of the Biomechanics and Biotransport Commons, Biomedical Devices and Instrumentation Commons, and the Data Science Commons


This Thesis (Senior Honors) is brought to you for free and open access by the Other Engineering Materials at Dartmouth Digital Commons. It has been accepted for inclusion in ENGS 88 Honors Thesis (AB Students) by an authorized administrator of Dartmouth Digital Commons. For more information, please contact dartmouthdigitalcommons@groups.dartmouth.edu.
UTILIZING NEURAL NETWORKS AND WEARABLES TO QUANTIFY HIP JOINT ANGLES AND MOMENTS DURING WALKING AND STAIR ASCENT

by

MEGAN MCCABE

Bachelor of Arts Honors Thesis

Thayer School of Engineering
Dartmouth College
Hanover, New Hampshire

Date June 6th, 2020

Approved: [Signature]
Advisor’s Signature

Megan McCabe
Signature of Author
Abstract

Quantifying hip joint moments is critical to analyzing failure mechanisms and improving designs of total hip arthroplasty (THA) implants, which impact the health of millions of patients across the United States. The gold standard for computing hip joint angles and moments relies on optical motion capture and force plates, which are expensive and non-portable. This study developed two, more portable approaches for analyzing walking and stair ascent in the sagittal and frontal planes. The Insole-Standard (I-S) approach replaced force plates with force-measuring insoles, allowing for many gait cycles to be captured in succession on a treadmill and stair exercise machine. I-S results matched the curvature of results from similar studies, but peak kinetic results were high due to error induced by applying the vertical ground reaction force to the talus rather than modeling movement of the application point. I-S stair ascent results exhibited a peak flexion moment that is not found in the curvature of results from similar work, which may be partly attributed to moving steps on a stair exercise machine. The Wearable-ANN (W-A) approach combined the insoles with inertial measurement units and artificial neural networks (ANN) to compute the same results. A simple ANN with two hidden layers, five nodes in each, performed best. Compared against I-S results, the W-A approach performs well (average rRMSE = 16%, $R^2 = 0.81$ across outputs, activities, and training rounds), demonstrating a simple approach (2-3 wearables, 10 hidden nodes) can estimate hip kinematics and kinetics in two planes with relatively high accuracy. Future work should characterize the sensitivity of the approach to the precision of syncing between sensing modalities and to the degree of variability within and between training and test datasets. Data augmentation or ANNs trained for specific subject groups (i.e. split by age, gender, and/or pathology) may improve
results. The W-A results in this study are promising and with further improvement of the technique, it could prove invaluable for characterizing THA patient kinematic and kinetic data in their home environments.
Acknowledgements

First and foremost I would like to thank everyone at the Dartmouth Biomedical Engineering Center (DBEC) for their support and mentorship over the last four years. Barbara Currier provided me mentorship during my first year as a Women In Science Program (WISP) intern. Her support helped me develop a passion for research that carried me through my four years. She graciously continued to offer her support after WISP, providing FTIR expertise for my ENGS 89-90 team. Professor Van Citters acted as my mentor from WISP through my senior honors thesis research. I will be forever grateful for the time and energy he has invested in my work and my development as a scholar. I admire his strong passion for research and teaching, and his drive to shape Thayer into the best engineering school that it can be. Ryan Chapman acted as my mentor during my junior and senior year conducting research at DBEC. I would like to thank him for pouring countless hours into meeting with me for check-ins and analysis, helping me obtain a lab space for my honors thesis research, and reading through many long drafts of this manuscript and others. Finally, I thank him for sharing his enthusiasm for biomechanics and data analysis with me. I will take this passion with me as I move on to the next stage of my life. Thank you to Kori Jevsevar, Mike Kokko, John Currier, Hannah Grover, Peder Solberg, Audrey Martin, Professor John Collier, and Carolina Lago Pena Maia for supporting my research and providing a friendly community at DBEC.

Thank you to all of the subjects who participated in this study. Due to COVID-19, I was pressed to complete data capture much earlier than expected. I am forever grateful to the individuals who stepped up to participate in my study at the very last minute.
I would also like to thank the Neukom Scholars Program and the Dartmouth Undergraduate Advising and Research Office for providing funding for the project. Thank you to WISP for starting my research journey.

Finally, thank you to my friends, family, and boyfriend, Chris, for your support in completing all of my research over the last four years. Chris, in particular, has stood by my side throughout all of my projects, often bringing me food when I was stuck in Thayer all day long and listening as I talked endlessly about interesting findings I had made or about frustrations I had coding, writing, or otherwise completing the work. His help recruiting subjects at the last hour of winter term made finishing this project possible.
Table of Contents

Abstract ........................................................................................................ iii
Acknowledgements ....................................................................................... v
List of Tables ................................................................................................. ix
List of Figures ............................................................................................... x
List of Acronyms ........................................................................................... xi

1. Introduction ................................................................................................. 1

2. Methods ...................................................................................................... 9
   2.1 Data Capture .......................................................................................... 9
   2.2 Data Pre-Processing ............................................................................. 13
      2.2.1 Overview ....................................................................................... 13
      2.2.2 MATLAB Pre-Processing ............................................................. 14
      2.2.3 Syncing ......................................................................................... 16
      2.2.4 OpenSim Workflow ..................................................................... 20
      2.2.5 ANN Workflow .......................................................................... 23
         2.2.5.1 ANN Transfer Functions ....................................................... 24
         2.2.5.2 Size of the ANN ................................................................. 25
         2.2.5.3 Training Algorithm and Parameters ..................................... 26
         2.2.5.4 ANN Design Workflow ...................................................... 29

3. Results ....................................................................................................... 30
   3.1 Insole-Standard Approach .................................................................. 30
   3.2 Stair Ascent on Exercise Machine ...................................................... 33
   3.3 Wearable-ANN Approach ................................................................... 33
3.3.1 ANN Design ................................................................. 33
3.3.2 ANN Performance ....................................................... 38

4. Discussion ........................................................................... 40

4.1 Insole-Standard Approach .................................................. 41
4.2 Stair Ascent on Exercise Machine ........................................ 45
4.3 Wearable-ANN Approach .................................................... 47
  4.3.1 ANN Design ................................................................. 47
  4.3.2 ANN Performance .......................................................... 50

4.4 Significance ........................................................................ 53

Appendices .............................................................................. 56

Appendix A: Diagram of Final ANN Architectures ....................... 56

Appendix B: Ensemble Averaged I-S Results Across Subjects .......... 57

Appendix C: Ensemble Averaged nGRFs Across Subjects .............. 57

References ............................................................................. 58
List of Tables

Table 1: Gait parameters and subject characteristics ........................................... 30
Table 2: Tabulated I-S peak moments compared to other studies ............................. 31
Table 3: Tabulated performance metrics for W-A approach ................................. 39
List of Figures

Figure 1: Study overview ................................................................. 8
Figure 2: Sensing modalities used in the study ................................. 10
Figure 3: Data pre-processing overview ........................................... 14
Figure 4: Loadsol data pre-processing correction for force measured when insoles were unloaded ......................................................... 15
Figure 5: Plots used to sync data between sensing modalities and compute gait percentage vector ................................................................. 16
Figure 6: Diagram showing the angle used to rotate the shank IMU data used to sync IMU data with MOCAP and Loadsol data ....................... 19
Figure 7: OpenSim workflow .............................................................. 20
Figure 8: Sigmoid function versus hyperbolic tangent sigmoid function ................................................................. 25
Figure 9: Ensemble averaged hip joint angles computed using the I-S approach .......... 32
Figure 10: Ensemble averaged hip joint moments computed using the I-S approach plotted against some example curves from the literature for comparison ...................... 33
Figure 11: Bar plots summarizing results from the ANN design workflow .......... 34
Figure 12: Ensemble averaged hip joint angles and moments computed using the I-S versus W-A approaches .......................................................... 38
Figure 13: Violin plots demonstrating W-A results by subject split by ANN output .... 40
Figure 14: Diagram demonstrating (A) the difference in the lever arm for a nGRF applied at the talus versus a nGRF applied at the true CoP and (B) the CoP trajectory during stance phase .................................................................. 43
# List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>Artificial Neural Network</td>
</tr>
<tr>
<td>BMI</td>
<td>Body Mass Index</td>
</tr>
<tr>
<td>CoP</td>
<td>Center of Pressure</td>
</tr>
<tr>
<td>CT</td>
<td>Computed Tomography</td>
</tr>
<tr>
<td>FFNN</td>
<td>Feed Forward Neural Network</td>
</tr>
<tr>
<td>FEM</td>
<td>Finite Element Models</td>
</tr>
<tr>
<td>GRF</td>
<td>Ground Reaction Forces</td>
</tr>
<tr>
<td>IMU</td>
<td>Inertial Measurement Unit</td>
</tr>
<tr>
<td>I-S Approach</td>
<td>Insole-Standard Approach</td>
</tr>
<tr>
<td>ID</td>
<td>Inverse Dynamics</td>
</tr>
<tr>
<td>IK</td>
<td>Inverse Kinematics</td>
</tr>
<tr>
<td>JRA</td>
<td>Joint Reaction Analysis</td>
</tr>
<tr>
<td>LOO-CV</td>
<td>Leave-One-Out Cross Validation</td>
</tr>
<tr>
<td>LSTM</td>
<td>Long Short Term Memory</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>nGRF</td>
<td>Normal Ground Reaction Forces</td>
</tr>
<tr>
<td>NHN</td>
<td>Number of Hidden Nodes</td>
</tr>
<tr>
<td>MOCAP</td>
<td>Optical Motion Capture</td>
</tr>
<tr>
<td>ROM</td>
<td>Range of Motion</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root Mean Square Error</td>
</tr>
<tr>
<td>rRMSE</td>
<td>Relative Root Mean Square Error</td>
</tr>
<tr>
<td>THS</td>
<td>Time of First Heel Strike</td>
</tr>
<tr>
<td>Total Hip Arthroplasty</td>
<td>THA</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Total Knee Arthroplasty</td>
<td>TKA</td>
</tr>
<tr>
<td>Wearable-ANN Approach</td>
<td>W-A Approach</td>
</tr>
</tbody>
</table>
1. Introduction

Net moments about the hip are commonly used to study overall hip joint load during walking and stair ascent, which are two of the most strenuous and common activities performed by total hip arthroplasty (THA) patients [1–3]. As such, characterizing net hip moments during these gait activities is critical to understanding many failure modes of THA, an increasingly common treatment for end-stage hip disorders, like osteoarthritis. THA is the second most common total joint arthroplasty procedure in the United States, with over 400,000 performed each year [4]. Wear, aseptic loosening, dislocation, and other mechanical failures account for over 80% of revision THA procedures reported to the American Joint Replacement Registry between 2012 and 2018 [5]. Net hip moments play a biological and mechanical role in these failure modes. Biologically, the torques imposed on the hip play a fundamental role in bone healing and growth, which are critical to bone-implant fixation and the overall health of the joint [2,6–8]. Mechanically, excessive or abnormal hip torques may cause implant loosening, increased wear, cracking, delamination, or other mechanical damage. The biological and mechanical consequences of excessive hip joint moments may explain higher rates of revision in THA patients who are obese, which have made up an increasing proportion of overall THA patients over time [9,10]. Computing net hip moments during walking and stair ascent informs simulation studies of the hip and THA implant models used for failure analysis.

Mechanical failure of THA implants is typically studied using in vitro laboratory simulations or finite element modeling (FEM). In vitro laboratory simulations utilize machines designed to enact physiologically relevant loads on models of the entire hip joint.
or a hip joint component (e.g. femur, acetabulum). These specimens are typically made up of synthetic or cadaveric bones and animal-derived joint serums (synthetic bones: [3,16], cadaveric bones: [17,18]). FEM takes anthropometric measurements and/or radiographs to emulate the hip joint [11–15]. The closeness of simulated loading to true in vivo loading of these models, both machine-based and FEM, determines their clinical value [3,16].

Directly measuring musculoskeletal loading requires instrumented implants, which are rare. Most studies use loading parameters determined through biomechanical simulations. These simulations compute joint kinematics (i.e. angles) as a preliminary step to computing kinetics (i.e. forces and torques), which can be used in THA failure studies.

OpenSim is one such biomechanics software, made available for free by Stanford University to allow greater accessibility and collaboration among biomechanists [19,20]. The software can be used to complete inverse kinematics (IK) to compute joint angles and segment positions, inverse dynamics (ID) to compute net joint moments, and joint reaction analysis (JRA) to compute joint contact forces. Most biomechanics studies compute net joint moments, which represent the sum of joint contact forces and muscle forces on the joint. Net joint moments are preferred because JRA requires optimizing individual muscle contributions, which is computationally complex, time-consuming, and requires expertise to achieve reliable results [21]. Further, JRA is highly sensitive to errors in the model used for computation, necessitating highly subject-specific models based off of medical imaging scans (i.e. MRI, CT, etc.) that limits subject enrollment in such studies [22–24]. In contrast, net joint torques are easily computed in OpenSim, providing valuable information about the total load applied at the joint during a studied activity.
The standard method for computing net joint moments using OpenSim or any other biomechanics software requires optical motion capture (MOCAP) and force plate data as inputs. OpenSim provides a tool for scaling the model to better match subject dimensions and compute kinematic and kinetic results using IK and ID respectively. MOCAP consists of a system of cameras that track the position of reflective markers fixed to subjects’ body segments. OpenSim IK takes MOCAP marker trajectories as inputs and adjusts model positioning to minimize the difference in model marker positions with actual marker positions. The kinematics outputs of IK consists of angles that are computed between segment axes of the corrected model. Force plates measure the ground reaction force (GRF), or the force of the ground on the foot, as well as the center of pressure (CoP), or the application point for the GRF. OpenSim ID takes the kinematic output of IK and the GRF as inputs to compute the net moments on model joints. The standard method has two critical shortcomings: 1) the technology is expensive (Force Plate >$10K [25], MOCAP fixed cost >$14K [26,27]), and 2) the approach is constrained to a laboratory space. These limitations put significant spatiotemporal constraints on data capture. For example, walking and stair ascent are typically captured by embedding a force plates into walkways or steps. Some researchers use multiple, adjacent embedded force plates or instrumented treadmills to record gait cycles in succession, but both options are costly. To reduce cost, many researchers use a single embedded force plate, which makes capturing gait cycles in succession impossible [28–30]. A portable alternative to data capture using MOCAP and force plates could make it possible to capture patient-specific kinematic and kinetic data in their home environments, including gait cycles in succession.
Wearable sensors, like inertial measurement units (IMUs) and force-measuring insoles, allow for portable capture of kinematic and GRF data. IMUs are small, electromechanical devices that are fixed to body segments (i.e. thigh, shank, etc.) to measure triaxial, segmental acceleration, angular velocity, and magnetic field strength. The orientation of the IMU in a global reference frame can be represented by a quaternion, which is a four element vector used to describe rotations in 3-D space. Force-measuring insoles are slipped into shoes to measure GRFs. Using force-measuring insoles for capture of GRFs allows for many gait cycles to be captured in succession and negates the issue of “targeting” induced by force plates, which is the concern that subject’s movement might be altered to “target” foot contact with the force plate during a study [31]. The use of force-measurement insoles and IMUs together represents a viable, portable alternative to MOCAP and force plates for capturing kinematic and GRF data in any environment, in and outside of the laboratory [32–34].

MOCAP markers are primarily placed on bony landmarks, capturing data in a laboratory coordinate system that is easily transferred to biomechanical models. In contrast, wearable sensors, like IMUs, are not always fixed to bony landmarks and they capture data in a local (i.e. device) reference frame. Combining wearable device data with traditional biomechanical modeling may require limiting the placement of wearables to bony landmarks so that the device could be defined in the coordinate system of particular segments in the model. Then complex coordinate transformations would need to be completed to convert wearable data from its local coordinate frame to the body segment’s,
and finally to the laboratory coordinate frame describing the overall model position. Some studies have computed joint kinematics and kinetics using wearables and biomechanical modeling, but computations are time-consuming and require many IMUs [35–37].

Machine learning can be leveraged to bypass the need to complete complex coordinate transformations of wearable data, reducing computation complexity and time. Artificial neural networks (ANN), one of the most common machine learning algorithms used in biomechanics [38], estimate nonlinear relationships between inputs and outputs, like those existing between segment kinematics and GRFs as inputs and joint angles and moments as outputs. ANNs emulate biological neurons, consisting of a series of computational “nodes” which take weighted sums of inputs and transform them using nonlinear activation functions. Training the algorithm consists of using a training data set with known joint angles and moments to optimize ANN weights and biases, the coefficients of node inputs used in weighted sums and constant terms added to nodes respectively. These weights, or matrices of coefficients, mimic state space equations used widely in engineering and mechanics problems, which may provide justification for replacing traditional mechanical methods with ANNs. Perhaps due to improvements in computing capacity, interest in using ANNs to compute joint kinematics and kinetics has grown only recently.

In two studies, Mundt et al. investigated the performance of feed forward neural networks (FFNN) and long short term memory cells (LSTM) for joint kinematic and kinetic computations [39,40]. FFNNs are the simplest form of neural networks, consisting of a series of layers that feed data from one layer to the next, in order and without any feedback.
LSTMs can be incorporated into FFNNs, creating feedback loops that are often used to “learn” order dependence in sequence prediction problems [41]. Mundt et al. achieved relatively accurate results for predicting lower body joint angles and torques during walking, with an average RMSE smaller than 4.8° for hip joint angles and an average relative RMSE (rRMSE: relative to average range of predicted moments and ground truth moments) smaller than 13% for hip joint moments across subjects. Although successful, this approach may be unnecessarily complex. For example, Mundt et al. created “simulated” IMU data by computing body segment accelerations and angular velocities from MOCAP data which were more readily available to them. This allowed them to bypass the need for collecting IMU data in the first study and to supplement their measured data from 23 subjects in the second study, but it also induced computational complexity to the approach without an established means of checking how “simulating” IMU data might induce error into the final results. Further, the neural networks consisted of thousands of neurons (4000-6000) per hidden layer and 12,500-15,000 training steps. In contrast, one group validated a much simpler approach, using a single IMU on the waist to predict GRFs, joint angles, and joint moments of the lower body in the sagittal plane. With data captured from seven subjects walking on a treadmill, they built a simple ANN (20 hidden nodes). This approach achieved a rRMSE of 3.14 ± 1.49° for computing hip joint angles and 10.74 ± 1.26% for computing hip joint moments in the sagittal plane, proving simpler ANNs are capable of accurately computing joint kinematics and kinetics. There is a need for investigating how simpler ANNs (<50 hidden nodes) might perform computing more complex outputs (i.e. across multiple activities and/or in different planes of motion).
This study seeks to develop a Wearable-ANN (W-A) method for computing sagittal and frontal hip joint angles and moments using IMU and force-measuring insole data as inputs into an ANN with a relatively simple architecture (1-2 hidden layers, <100 nodes/layer) (Figure 1: Bottom). Based on prior work combining wearable data with ANNs to compute joint angles and torques, I hypothesize that my W-A method will achieve an average rRMSE across subjects of less than 20% [39,40,42,43]. This goal rRMSE is higher than the Mundt and Lim studies because this study seeks to predict results in two planes (sagittal and frontal) with relatively small datasets and simple ANN architectures. The results of the W-A method will be compared with those of a quasi-standard approach, one consisting of the traditional inverse kinematics and dynamics workflow, but replacing force plates with force-measuring insoles (Figure 1: Top). Most studies seeking to replace force plates with portable alternatives for biomechanical modeling and inverse dynamics combine pressure-sensing insoles and custom developed algorithms that compute the center of pressure (CoP) [44,45]. Force-measuring insoles impose limitations on the approach because they only capture vertical GRFs, not anterior-posterior or medial-lateral forces, they do not capture CoP data, and they capture at a lower sampling frequency than do force plates. In particular, we hypothesize that our choice to apply the vertical GRF to the talus in our I-S approach will lead to higher maximum flexion moments based on a study by McCaw et al., which demonstrated that posteriorly shifting the CoP from its true position resulted in greater maximum flexor torque at the hip during gait [46]. Despite this, we also hypothesize that the overall curvature of the I-S computed hip angles and moments will mimic that of similar studies found in the literature using a fully gold-standard approach, achieving maximum flexion moments within 25% of those computed in a previous study with university aged
participants and completing both walking and stair ascent (1.13 Nm/kg for walking, 0.80 Nm/kg for stair ascent; [2,28–30,47–53]). We expect hip angles computed in this study to match well with the literature since those computations do not rely on data from the force-measuring insoles.

The study will analyze two movements, stair ascent and walking, which represent two of the most common and relatively strenuous activities performed by THA patients [54]. Both walking and stair ascent are gait activities, consisting of a cycle that may be identified as starting at heel strike or toe off and ending at the next. The cycle may be divided into the stance phase, in which the foot is in contact with the ground, and the swing phase, in which the foot is held in the air to take a step. Many walking and stair ascent studies focus on computing joint kinematics and kinetics in the sagittal plane because it aligns with the direction of motion, but prior work shows that THA patients may exhibit greater biomechanical deviations from healthy subjects in the frontal plane due to weaker abductor muscles [55]. As such, this study will compute hip joint angles and moments in both the
sagittal and frontal planes. Ultimately, the study seeks to develop two alternative approaches to conventional biomechanical modeling for computing hip joint angles and net moments. The first approach relieves limitations imposed by force plates, such as the need for constructing and using specialized equipment (i.e. force plate-embedded walkways or steps) that constrains capture to limited spaces for short periods of time. The second approach offers a fully portable alternative to conventional methods, which if successful could allow for the study of THA patient biomechanics in their home environments for longer periods of time (i.e. hours or days as opposed to minutes captured in a laboratory). Improved knowledge of THA patient biomechanics following their procedure could allow for improved THA failure analysis and implant design. Implant design determines THA outcomes, which are critical to the health of millions of patients across the United States [56].

2. Methods

2.1 Data Capture

Broadly, data capture consisted of having subjects fill out a survey to determine foot dominance, measuring and recording subjects’ height and weight, fitting them with sensors, completing sensor calibration procedures, and collecting data as the subjects performed a set of walking and stair ascent trials. 17 Subjects (10M, 7F; average age 26.8 ± 6.4 years) were recruited from the university population following approval of the Institutional Review Board. Inclusion criteria were age ≥18 years, no musculoskeletal or neuromuscular impairments impacting the lower extremity, no terminal illness resulting in
death within one year, clinical full hip extension (≥ 10°) and flexion (≥ 100°) and complete participation in the study [57].

At the start of each session, subjects filled out a survey to determine foot dominance. The Waterloo Footedness Questionnaire lists a set of tasks (i.e. kick a ball, stand one one foot, etc.) and asks subjects to pick whether they would always or usually prefer one foot over the other [58]. All subjects expressed a preference for either the left or right foot (13 right dominant). Next, subjects’ height (1.74 ± 0.08 m), weight (81.6 ± 19.5 kg), and the angle between the lateral aspect or the shank and the medial surface of the tibia (121.3° ± 6.6) were measured and recorded. Subjects were then ready to be fitted with sensors (Figure 2).

![Figure 2: Sensing modalities used in this study: (A) MOCAP marker set (Helen Hayes Lower Body: 19 markers) (B) Left: APDM Opal IMU, Right: IMU fixation on the thigh and shank. (C) Loadsol iPad application and insoles [34].](image)

Data capture required three sensing modalities: MOCAP, force-measuring insoles, and IMUs. A system of six S250e cameras (OptiTrack Motive Body 1.10, NaturalPoint, Inc., Corvallis, OR) was calibrated according to manufacturer’s recommendations and used to track a modified lower body Helen Hayes marker set (19 markers; Figure 2A). Lower body bony landmarks (i.e. anterior superior iliac spine, sacrum, medial and lateral femoral
epicondyles, medial and lateral malleoli, between the distal ends of the 4th and 5th metatarsal, calcaneul tuberosity) were palpated and markers were adhered using double sided tape. One marker was also adhered to the lateral thigh and to the lateral shank. These markers were placed asymmetrically to allow the MOCAP cameras to distinguish between left and right legs. Most kinematic studies analyzing lower body motion use some modification of the Helen Hayes marker set because it is easy to implement, requiring relatively few markers which are placed primarily on bony landmarks [59,60]. Further, results may be more easily compared across studies that use similar marker sets to capture MOCAP data. Subjects were first fitted with MOCAP markers before donning IMUs.

The number and placement of IMUs varies across biomechanics studies, but the thigh and shank are commonly used to study walking and stair ascent [61–64]. It is also common for studies to only fix IMUs to the dominant leg. Computing joint torques of solely the dominant leg negates the need for more IMUs, simplifying the method and reducing the risk that fixing too many sensors to the subject will significantly affect their movement during capture. For this study, subjects donned two IMUs (APDM, Inc.; Portland, OR; f_s=128Hz), which were strapped using Velcro bands to the lateral aspect of the dominant thigh and the anteromedial aspect of the dominant shank (Figure 2B). The IMUs were calibrated per manufacturer’s instructions and set to log data continuously while undocked from the charging station. Data were later exported from the device SD cards to .apdm file format, converted to .h5 and then .mat files, and synced between units using their recorded epoch time stamps.
The last sensing modality used were force-measuring insoles. Novel Inc. offers low-cost force-measuring insoles, called “Loadsols” (Novel Electronics, St. Paul, MN, USA; [65]; Figure 2C). These insoles are slipped inside subjects’ shoes to measure the GRF normal to the surface (nGRF) of the insole during data capture. Prior work has validated Loadsols’ measurement of the nGRF against force plates for slow-to-moderate speed activities, such as walking [33,34,66–70]. In this study, stair ascent was completed at a slower speed than walking, and can also be considered a low speed movement. The Loadsol’s relatively low sampling frequency (100 Hz v. >1000 Hz for a typical force plate) may not be adequate for capturing fast movements, like running [34], but it should not be a significant limitation for capturing the movements in this study. Following IMU donning, subjects were fitted with Loadsol insoles, which were connected to the Loadsol app on an iPad (Apple, Cupertino, CA, USA) via Bluetooth for data logging. Subject mass was entered into the app prior to Loadsol calibration for each capture. The insoles captured at a frequency of 100 Hz and a working range of 20-2000N.

Once subjects were fitted with sensors, they stood still for 10 seconds for a ‘standing trial’, which was used to calibrate the MOCAP system and later used for OpenSim model scaling (See section 2.2.4). Then subjects completed three 30 second trials of walking on a treadmill (2 mph) and three 10 second trials of stair ascent on a stair-climbing exercise machine (StairMaster StepMill 7000PT; Speed level 8; 20.32 cm rise x 23.5 cm run). The activities were captured on a treadmill and stair exercise machine to allow for the capture of many gait cycles within the laboratory space (average gait cycles captured per subject per trial: 22.18 ± 1.70 (walking), 4.53 ± 0.62 (stair ascent)). Stair ascent trials were kept
short to prevent subject fatigue which has been shown to alter gait performance [71,72]. This was also important for maintaining subject safety. Data capture from each of the three sensing modalities were roughly synced by: 1) Starting capture for Loadsol and MOCAP recording for each trial at the same time, 2) using a fourth IMU and the data marking button made available by APDM to mark the start of each trial in the logged APDM data. To achieve more precise syncing, gait analysis was used during pre-processing to identify the time of initial heel strike in each type of data (MOCAP, Loadsol, IMU) and use it to temporally align data across sensors (See Section 2.2.3 Syncing).

2.2 Data Pre-Processing

2.2.1 Overview

Data pre-processing consisted of preparing raw, captured data for two workflows: 1) the OpenSim workflow to complete the I-S approach, and 2) the MATLAB ANN workflow to complete the W-A approach (Figure 3). Only one of the three trials captured were pre-processed for each activity. The MOCAP file with the least gaps in marker trajectories was used to determine which trial to pre-process and analyze. Raw data from each of the sensing modalities, Loadsol, MOCAP, and APDM IMUs, were filtered, synced using gait cycle analysis, and exported or formatted for the two workflows. Custom MATLAB scripts were written to complete pre-processing and the ANN workflow (Code Appendices Volume). Finally, further scripts were used to consolidate, plot, and compare results between approaches.
2.2.2 MATLAB Pre-Processing

Raw Loadsol (nGRFs) and APDM (IMU kinematic data) were imported into MATLAB for pre-processing. MOCAP data (marker trajectories) were edited using the Optitrack Motive software to interpolate and fill gaps in the trajectories as well as filter the data (low-pass Butterworth, $f_{\text{cutoff}} = 6$ Hz) before import into MATLAB. The MOCAP trial with the least gaps determined which of the three trials captured were to be fully pre-processed. Custom MATLAB scripts were written to remove gaps in and filter the Loadsol and APDM data with low-pass Butterworth filters ($f_{\text{cutoff}} = 10$ Hz for both data types) [34,43]. APDM data for each trial were separated in MATLAB using marks created by depressing a button attached to the fourth IMU (‘IMU marker truths’) held by the researcher during data capture. Occasionally, depressing the button resulted in multiple marker truths in rapid succession, which was likely the result of holding the button down for too long. For these cases, the first marker truth was used and the rest were ignored. The screened marker truths were used by a MATLAB script to split the data by trial, and separate out the files chosen for full pre-processing.
Loadsol pre-processing required resampling (100 Hz to 128 Hz, the sampling frequency of the MOCAP and IMU data) before syncing with data from other sensing modalities. For some subjects, one additional step of Loadsol pre-processing was required because the Loadsols would report a force even when the foot was unloaded (i.e. swing stage of gait cycle). This force likely represented pressure between the insole, foot, and shoe that should not be included as part of the recorded nGRF, which should only consist of force passing from the ground to the foot. As such, Loadsol data were plotted to identify subjects that required a vertical shift to ensure that the recorded force during the unloaded stage of the gait cycle was equal to zero. Then, one unloaded portion of the gait cycle was identified for the foot of interest for that subject and activity and the force averaged to determine the magnitude of the vertical shift. This shift was then applied to all of the data for that foot from the particular file. Finally, the results of this shift were compared to the foot that did not require a vertical shift to check for alignment (Figure 4).

**Figure 4:** Plot of nGRF of subject requiring a vertical shift during pre-processing to account for pressure between the insole, foot, and shoe during non-loaded portions of the gait cycle. In this case, data from the right foot was shifted, the original curve shown in red and the corrected curve shown in blue. The data from the left foot (no correction made) is shown for reference in black.
2.2.3 Syncing

Many researchers use MOCAP-force plate systems that sync between sensing modalities automatically, but this study used a novel combination of technologies that required implementing steps to ensure synchronization across sensor types. Syncing between sensors within the same system (i.e. right/left insole for Loadsol, and between thigh/shank IMU for APDM) was done automatically by the Loadsol software and was completed during import into MATLAB using epoch time for APDM data. Syncing across sensing modalities was completed in two steps. During data capture, the start of recording MOCAP and Loadsol data were manually synced (i.e. buttons were pushed simultaneously) with each other and the “marking” of logged APDM data. During data pre-processing, gait analysis was used to identify the time of first heel strike, termed “T\textsubscript{HS}”, in each of the three types of data, data were shifted to align at T\textsubscript{HS} of the MOCAP data, and files were shortened on each side of T\textsubscript{HS} to ensure consistent length of data between sensors. T\textsubscript{HS} was identified in each of the three types of data using examples from the literature (Figure 5; [73–76]).

![Figure 5: Plots used to sync data from different sensing modalities based on time of first heel strike (T\textsubscript{HS}: black Xs) and convert time vector to percent of gait cycle; (A) Plots used to sync data from one subject’s...](image-url)
walking trial, where the top plot is MOCAP heel marker (anteroposterior ('z') position and velocity) data, the middle is the nGRF, and the bottom is shank IMU angular velocity data. T\(_{HS}\) in the APDM and Loadsol data were shifted to align with the T\(_{HS}\) of the MOCAP data (indicated by grey arrow). (B) Syncing for stair ascent was identical to that for walking except the local minima in the vertical ('y') velocity of the toe following swing phase was used in the place of the other two metrics [74]. (C) Plot of Loadsol data for stair ascent trial used to check algorithm used to convert time to gait percentage; pink circles indicate heel strike and black circles indicate toe off.

For MOCAP data, T\(_{HS}\) was identified following the example of Zeni et al. for walking data and Foster et al. for stair ascent data [73,74]. Zeni et al. found that two times match well with T\(_{HS}\) for walking: 1) the time at which there is maximal distance between the position of the heel and sacral ('waistback') marker in the direction of motion (z-axis) and 2) the time at which the z-axis velocity of the heel passes through zero. Both approaches worked well, with the average difference between T\(_{HS}\) determined using a force plate and the T\(_{HS}\) determined using the proposed approaches being less than one frame (0.0167 sec). Both metrics were plotted for syncing (Figure 5A), but priority was given to aligning with the velocity metric to determine T\(_{HS}\) in MOCAP data because it showed slightly better results in the study by Zeni et al. [73]. Foster et al. validated a similar approach to determine T\(_{HS}\) in stair ascent data, showing that the vertical toe velocity (y-axis) exhibits a local minima just following swing phase that aligns well with T\(_{HS}\) (within an average of 0.040 sec). The y-axis toe velocity was plotted to sync stair-ascent data (Figure 5B).

Loadsol data were plotted with a line indicating a 20N threshold, the low end of the Loadsol recording range (Figure 5A, middle plot). T\(_{HS}\) can be clearly identified as a point where the data crosses the 20N threshold with a positive slope, leading to two local maxima characterizing stance phase. The approach of using the low-end of the force recording range for heel strikes identification is common among other gait event identification studies [73,74,77]. Toe off was identified similarly as the points where data passed the 20N threshold.
threshold with a negative slope. Toe off was identified so that the average percent stance duration per cycle could be computed and used to compare findings in this study to others analyzing gait.

Finally, APDM data were synced using the angular velocity of the shank about the axis running mediolaterally, perpendicular to the direction of motion, similar to that used in other studies identifying $T_{HS}$ for both stair ascent and walking [75,76,78]. Typically studies using a shank mounted IMU to identify heel strike fix the IMU to the lateral aspect. This maximizes rotation about the mediolateral axis during gait. This study fixed the IMU to the anteromedial aspect of the shank, or the medial surface of tibia, to reduce soft tissue noise. To account for the difference in IMU positioning on the shank, this study rotated shank data used for syncing (i.e. shank angular velocity data was copied to allow for processing on one copy for syncing and separate processing on the other copy for further calculations). The rotation was completed by multiplying data by a rotation matrix that utilized an angle, $\alpha$, measured during data capture between the medial surface of the tibia and the lateral aspect of the shank. Figure 6 demonstrates the relative placement of the shank IMU in this study to that used in literature studies as well as the angle of rotation ($\alpha$) in the transverse plane cutting through the shank.
This angle was approximated during data capture using a goniometer to measure the angle between the axis normal to the face of the IMU against the shin, pointing outwards and the mediolateral axis (Figure 6B; \( \alpha = 121.3 \pm 6.6^\circ \)). The shank syncing data was also filtered at a lower cutoff frequency than that of the rest of the APDM data (\( f_{\text{cutoff}} = 2.3 \) Hz v. 10 Hz for non-syncing APDM data) following the example of Yang et al. [78]. Finally, the rotated, filtered shank angular velocity data was plotted to identify local minima following a prominent local maxima, corresponding with T\(_{\text{HS}}\) and swing phase respectively (Figure 5A, bottom plot).

After all data were synced across sensors, a common time vector was used to calculate a gait percentage vector, a vector which started at 0% at every T\(_{\text{HS}}\) and progressed to reach 100% just before the next T\(_{\text{HS}}\). Heel strike and toe off were identified in Loadsol data, as previously described, and used to compute the gait percentage vector. Results were checked by plotting (Figure 5C) and when the plot showed erroneous points of heel strike.
or toe off, the false points were removed manually before the gait percentage vector was calculated.

2.2.4 OpenSim Workflow

OpenSim, the opensource biomechanics software from Stanford, was used to calculate hip joint kinematics and kinetics for the Insole-Standard method using MOCAP and Loadsol data [19,20]. This workflow consists of three main steps: Scaling, Inverse Kinematics (IK), and Inverse Dynamics (ID), which are summarized in Figure 7. Settings for each step were based off of OpenSim Tutorial 3 [79]. Modifications to these settings are mentioned as they relate to the three workflow steps discussed below.

**Figure 7**: OpenSim workflow listing inputs and outputs (and file formats) required for each of the three steps: scaling, inverse kinematics, and inverse dynamics.

First, the scaling tool was used to fit the generic gait2392 model available in OpenSim to each subject. The gait2392 model is a 3-dimensional, 23 degree-of-freedom linked-segment model of the human musculoskeletal system including 92 musculotendon actuators that represent 76 muscles in the lower extremities and torso [80]. The default 1.8 m, 75.16 kg model is scaled to fit subject-specific dimensions by comparing the distance between specific markers in the experimental file ($e_1$) to the same distance in the existing
model \((m_1)\) for a standing trial captured at the beginning of each data capture session. For instance, the pelvis is scaled by computing the scaling factor, \(s\), that corresponds to the distance between the left and right anterior superior iliac spine (L.ASIS, R.ASIS respectively) in the experimental file to that in the existing model \((s = e_1/m_1)\). If more than one measurement describes the model segment (i.e. L.ASIS to R.ASIS and L.ASIS to sacrum), the overall scaling factor is an average of the scaling factors computed from each measurement [81]. This study used a standing trial as the experimental file for scaling as is typical. Pre-processed MOCAP data (marker trajectories) from a standing trial were formatted in a .trc file using a custom MATLAB script for import into the OpenSim scaling tool. This study manually scaled the height of the model using the measured height of the subject because no MOCAP markers were placed on the upper body.

The subject-specific scaled model was then combined with MOCAP data from the stair and walking trials to compute joint kinematics using inverse kinematics (IK). Following MATLAB pre-processing, MOCAP data were exported in .trc file format using a custom MATLAB script for import into OpenSim. During IK, OpenSim was used to solve the weighted least squares equation (Equation 1) to minimize error between modeled marker trajectories and experimental ones [82]. MOCAP markers placed on bony landmarks were given ten times higher weights than those placed on soft tissue (i.e. thigh and shank markers).

**Equation 1** \[
\min \left[ \sum_{i \in \text{markers}} w_i \left\| \mathbf{x}_i^{\text{exp}} - \mathbf{x}_i(q) \right\|^2 \right]
\]

Where \(i\) corresponds with the marker of interest, \(\mathbf{x}_i^{\text{exp}}\) with the position of the experimental marker, and \(\mathbf{x}_i(q)\) with the position of the model marker which is a function of the position of the body segment described by the vector \(q\).
The resulting model marker trajectories were used to calculate lower body joint angles as well as the orientation and position of the pelvis for each frame of the imported motion file. IK results are exported from OpenSim in .mot format. Only hip joint angles were analyzed, but all IK results were used as inputs into the OpenSim inverse dynamics (ID) tool.

OpenSim ID takes IK results, which fully define the motion of the model, to compute the system mass matrix, Coriolis and centrifugal forces, and gravitational forces. By summing these forces and the GRFs (Equation 2, Left side), the net joint moment can be computed (Equation 2, Right side; [83]).

\[
M(q)\ddot{q} + C(q, \dot{q}) + G(q) + F_{external} = \tau
\]

\(q, \dot{q}, \ddot{q} \in \mathbb{R}^N\) are the vectors describing the positions, velocities, and accelerations of the model respectively, \(M(q) \in \mathbb{R}^{N \times N}\) is the system mass matrix, \(C(q, \dot{q}) \in \mathbb{R}^N\) is the vector of Coriolis and centrifugal forces, \(G(q) \in \mathbb{R}^N\) is the vector of gravitational forces, \(F_{external}\) is the GRF, and \(\tau \in \mathbb{R}^N\) is the vector of model forces, and \(N\) is the number of degrees of freedom.

OpenSim requires GRFs be stored in .mot format for ID, with a corresponding .xml header file. In this study, a custom MATLAB script was written to export pre-processed Loadsol nGRF data into the .mot file format. The .mot file was written to apply the nGRF to the approximate center of the talus, at the midpoint between the MOCAP ankle markers, following the example of Dudam et al., who used a similar approach to apply normal GRFs captured by instrumented pedals on a stationary bike for inverse dynamic analysis [84]. A .xml header file was created for each of the .mot GRF files by copying .xml header files from the same OpenSim tutorial mentioned previously [79]. The only change made to these .xml files was to update the filepaths to point to the .mot nGRF files created in this study.
Finally, IK and ID results (hip joint angles in .mot format and moments in .sto format respectively) were exported from OpenSim and imported into MATLAB using custom scripts. Data were tabulated, and ensemble averaged. Ensemble averaging consisted of splitting data within each subject by gait cycle using the gait percentage vector. Next, data were resampled to attain 1000 points of data between 0 and 100% for each cycle. Gait cycles of each activity could then be aggregated across subjects and averaged to attain one angle and one moment curve for the sagittal plane and frontal plane each.

2.2.5 ANN Workflow

For the W-A approach, this study developed a shallow (i.e. <3 hidden layers) feed forward neural network (FFNN) to estimate the relationship between wearable data as inputs and hip joint angles and moments as outputs. This application would fall under the class of function approximation, where ANNs are used to model the relationship between continuous variable inputs and outputs. The inputs in this study include the duration of the activity (‘time’), the normal GRF (dominant foot only), and the IMU data from the shank and thigh sensor. Each IMU represents 13 inputs to the ANN (acceleration, angular velocity, and magnetic field of the thigh/shank x 3 dimensions + the orientation as a quaternion x 4 dimensions = 13 total variables/segment), which makes 28 total possible inputs to the ANN (time + nGRF + 13 thigh IMU variables + 13 shank IMU variables). The outputs of the ANN were the hip joint angles and moments in the sagittal and frontal planes (i.e. flexion angle, adduction angle, flexion moment, and adduction moment). For an ANN with only one hidden layer, prediction works as follows. Each node in the first layer of an ANN takes a weighted sum of the inputs and transforms the sum using a transfer
function. A weighted sum of the resulting values is then passed to the final output layer and another transfer function is applied to reach the final results. Training consists of tuning the weights between the layers, and terms called biases, which are constants added to each layer’s weighted sums. Designing the ANN required choosing architecture and training parameters. The architecture, or building blocks, of a shallow ANN include the number of hidden layers (1-2), the number of nodes in each layer, and the transfer functions applied to the nodes in the hidden and output layers.

2.2.5.1 ANN Transfer Functions
Variations of the sigmoid function are commonly used in hidden layers and linear functions are commonly used in output layers for ANN function approximation [42,43,85]. Sigmoid functions map inputs to values between zero and one. Their derivative is easily computed (Figure 8; Top equations), making ANN training more efficient. Hyperbolic tangent sigmoid (“tansig”) transfer functions offer further advantages over the sigmoid function due to their greater slope and range ([-1,1] versus [0,1] for original sigmoid; Figure 8). The function’s steep slope magnifies small changes in the input variable. Further, the function exhibits an operating range that spans negative and positive numbers [86]. Following the example of Stetter et al., this study used a hyperbolic tangent sigmoid function for hidden layers and a linear function for the output layers [87].
2.2.5.2 Size of the ANN

ANN capability in modeling complex, nonlinear relationships increases with increasing number of hidden nodes (NHNs). However, too many nodes may allow the ANN to overfit to the training data, making them poorly suited for predicting relationships in new data sets. Further, the larger the ANN, the longer the computation time for training. Researchers using ANNs to predict joint kinematics and kinetics typically look to prior work using ANNs for biomechanical applications as starting points for their own ANN architectures [39,42,43]. Then, they use trial and error, training ANNs of slightly different sizes to determine which one performs best with the dataset of interest. Lim et al. captured data from 7 subjects and proposed an ANN with one hidden layer of 20 nodes [43]. Stetter captured data from 13 subjects and proposed an ANN with two hidden layers of sizes 100 and 25 nodes [42]. Based on the example of Lim et al. and Stetter et al., this study had 17 subjects and chose an ANN with 25 NHN in the first layer and 10 in the second layer as a
starting point to iterate using a ANN design workflow. Hereafter, ANNs with two hidden layers will be referred to as [NHN in first layer, NHN in second layer] for simplicity (i.e. ‘[25,10]’ for the starting point ANN).

The ANN design workflow consisted of training and comparing ANNs of difference sizes (i.e. different number of nodes and hidden layers). ANN performance is commonly tracked using pearson’s coefficients and relative root mean square error (rRMSE; Equation 3). However, Pearson’s coefficients require subjective classification by study authors as ‘strong’, ‘moderately strong’, or ‘weak’. In contrast, a coefficient of determination ($R^2$) close to one clearly signifies that a predicted variable follows the curvature of a ‘ground truth’ variable. This study used rRMSE, and replaced Pearson’s coefficients with the coefficient of determination ($R^2$) which is more easily interpreted.

\[
\text{Equation 3} \quad \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \epsilon^2} \quad \text{rRMSE} = \frac{\text{RMSE}}{\frac{1}{2}[(p_{max} - p_{min}) + (o_{max} - o_{min})]} \times 100\%
\]

Where $n$ is the number of data points, $\epsilon$ is the error (predicted or W-A computed value – observed or I-S computed value), $p_{max}, p_{min}$ are the maximum and minimum predicted values, and $o_{max}, o_{min}$ are the maximum and minimum observed values.

2.2.5.3 Training Algorithm and Parameters

Training a FFNN consists of modifying weights and biases in the ANN to achieve optimal performance. In biomechanics, subjects’ data are typically used for either training (‘training set’) or validation (‘validation set’) to avoid bias [38]. Leave-one-out cross validation (LOO-CV) is commonly used in biomechanics and consists of training an ANN with all subjects’ data but one, which is left as the validation dataset. Training is repeated
to allow each subject a chance to be the validation set and performance metrics are averaged from the separately trained and tested ANNs.

ANN training can be split into four stages: 1) initialization, 2) a forward pass, 3) computation of the loss function, and 4) a backwards pass [88]. Initialization consists of filling the weights and biases with initial values. This study uses the Nguyen and Widrow algorithm available as “initnw” in MATLAB, which initializes weights and biases so that they randomly cover the active range of the input space, making training more efficient. Inputs were normalized to a range of [-1,1] to allow for more efficient and effective training in this stage as well (mapminmax function in MATLAB). Initialization is followed by a forward pass, where the ANN computes an output from the initial weights and biases and inputs. This output is compared to the ‘ground truth’, or the response variable of the training dataset. The loss function is computed, relating the error between the model’s current output and the given ‘ground truth’ to the weights and biases used for that forward pass. Next, a backwards pass propagates the error back through the layers of the ANN in a process called backpropagation. The contribution of each weight to the total error can be discerned and the weight updated using various techniques, but the most common is gradient descent. The gradient is taken of the loss function with respect to each weight and used to update the weight and improve performance. This approach is often compared to a ball in a 2-D bowl. If the slope is highly positive, then moving the ball far to the left would move it toward the local minima in error. If the slope is slightly negative, moving the ball slightly to the right would move it toward the minimum error. In reality, the loss function is multivariable and the gradient is used to find the global minima.
This study used the Levenberg-Marquardt algorithm in MATLAB ("trainlm") for training. It is MATLAB’s most efficient ANN training algorithms for shallow networks, employing two main techniques to achieve efficiency. First, the algorithm approximates the Hessian matrix typically used in gradient descent as $H = J^T J$, computing the gradient as $g = J^T e$, where $J$ is the Jacobian matrix and $e$ is the vector of network errors. The Hessian is a square matrix of second order partial derivatives of a scalar-valued function. The matrix describes the curvature of a function of many variables, but is difficult to compute. The Jacobian is a matrix containing the first order partial derivatives of a vector valued function. It is easier to compute and can be used to approximate the curvature of a multivariable function like the loss function of an ANN.

The second element that allows for the LM algorithm’s efficiency is the use of a learning parameter, $\mu$, which selectively emphasizes certain terms in the algorithm as training progresses. Equation 4 shows that the size of $\mu$ determines the importance of the Hessian approximation, $J^T J$ in computing the new value for the weight. This parameter starts small, which makes the LM algorithm act like a gradient descent algorithm with a small step size. Later in training, the parameter is decreased, allowing the LM algorithm to approximate Newton’s Method, which is more efficient and accurate near a local minimum.

**Equation 4**  
\[ x_{k+1} = x_k - [J^T J + \mu I]^{-1} J^T e \]  
[89]

*Where $x$ is the network weight or bias term, $k$ is the training iteration, $\mu$ is the learning parameter, $J$ is the Jacobian matrix containing first derivatives of network errors with respect to the weights and biases, and $e$ is the vector of network errors*
None of the default parameters for `trainlm` were changed for this study. The maximum number of epochs, or full passes through the training data, was kept at 1000 and training was ended early if the gradient stopped decreasing for more than six epochs. In this study, training never progressed further than 100 epochs.

**2.2.5.4 ANN Design Workflow**

In overview, the ANN design workflow used in this study was: 1) train an ANN of the same architecture 10 times (10 rounds), 2) average the rRMSE within rounds (across LOO-CV iterations) but keep metrics for each output (i.e. flexion angle, adduction angle, etc.) separate, 3) use the Anderson-Darling test to determine whether rRMSE results across testing rounds were approximately normally distributed, 4) Use two-tailed t-tests (alpha = 0.05) to detect significant differences in rRMSE between selected ANN architectures.

Initialization determines how quickly and accurately ANN training algorithms converge [88]. Different initial weights and biases lead to slightly different final solutions. As such, variation due to initialization had to be characterized before investigating the effects of different ANN architectures on performance. One round of ANN training consists of 17 LOO-CV training events, one event per subject acting as the validation set. Ten rounds of training were completed for each ANN architecture.

Next, rRMSE for each output was averaged across LOO-CV iterations leaving 10 sample points (one for each round) per 4 sample sets (one for each ANN output). Each sample set
was tested using the Anderson Darling test (‘adtest’ in MATLAB, $\alpha = 0.05$) to ensure they could be approximated as coming from a normally distributed population.

Finally, a two-tailed t-test with $\alpha = 0.05$ was used to distinguish between the results of one ANN architecture versus another. ANN iterations differed by the size of the input and hidden layers. ANN design consisted of three stages to investigate, 1) the size of the first hidden layer, 2) the size of the second hidden layer, and 3) the size of the input layer. Design prioritized performance (low rRMSE across outputs), then simplicity (fewer nodes), and finally consistency between the ANN used with walking data and the ANN used with stair ascent data.

3. Results

3.1 Insole-Standard Approach

The gait parameters and subject characteristics are summarized below in Table 1. Subject mass had a higher standard deviation than that found in other similar studies (SD = 19.5 kg (this study), 7.7 kg [43], 12.7 kg [40]). Stance phase made up approximately 66% of the gait cycle for both walking and stair ascent. Stride durations were shorter for walking than for stair ascent.

Table 1: Gait parameters and subject characteristics.

<table>
<thead>
<tr>
<th>A. Gait Parameters</th>
<th>B. Subject Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stride Duration (s)</td>
<td>Height (m)</td>
</tr>
<tr>
<td>Walking</td>
<td>Stair Ascent</td>
</tr>
<tr>
<td>% Stance</td>
<td>% Swing</td>
</tr>
</tbody>
</table>

Tabulated peak flexion, extension, and abduction moments are summarized in Table 2 below. Overall, I-S computed peak moments are larger than those computed in the
literature using a fully gold standard approach. I-S peak flexion and extension moments computed for walking are roughly twice those reported by Eng and Winter [30]. For stair ascent, most studies report a lack of a peak flexion moment, but the I-S approach exhibits a flexion moment of similar size the the extension moment. The I-S computed extension moment is the closer to that reported by Costigan et al. than walking moment peaks are to respective literature comparisons, but the difference is still notable [29].

<table>
<thead>
<tr>
<th>(A) Walking Peak Moments (N·m/kg)</th>
<th>Flexion</th>
<th>Extension</th>
<th>Abduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-S Results</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Costigan et. al.</td>
<td>2.15 ± 0.55</td>
<td>2.43 ± 0.6</td>
<td>1.62 ± 0.32</td>
</tr>
<tr>
<td>Eng &amp; Winter</td>
<td>-</td>
<td>1.13 ± 0.30</td>
<td>0.95 ± 0.14</td>
</tr>
<tr>
<td>Hunt et. al.</td>
<td>1.0 ± 0.30*</td>
<td>1.3 ± 0.35*</td>
<td>1.2 ± 0.25*</td>
</tr>
<tr>
<td>Pizzolato et. al.</td>
<td>0.56 ± 0.16</td>
<td>0.58 ± 0.60</td>
<td>0.80 ± 0.14</td>
</tr>
<tr>
<td>Rienert et. al.</td>
<td>0.3*</td>
<td>0.67*</td>
<td>0.69*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(B) Stair Ascent Peak Moments (N·m/kg)</td>
<td>Flexion</td>
<td>Extension</td>
<td>Abduction</td>
</tr>
<tr>
<td>I-S Results</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Costigan et. al.</td>
<td>1.05 ± 0.26</td>
<td>1.41 ± 0.21</td>
<td>1.33 ± 0.22</td>
</tr>
<tr>
<td>Rienert et. al.</td>
<td>0.1*</td>
<td>0.80 ± 0.24</td>
<td>0.80 ± 0.24</td>
</tr>
<tr>
<td>Protopapadaki et. al.</td>
<td>0.16*</td>
<td>0.54*</td>
<td>-</td>
</tr>
<tr>
<td>McFayden &amp; Winter</td>
<td>0.33*</td>
<td>1*</td>
<td>3*</td>
</tr>
</tbody>
</table>

*Table 2*: Tabulated peak moments normalized to subject mass computed using the I-S approach. Adduction reaches a roughly flat plateau at zero during the swing phase of gait, but does not exhibit a sharp peak, hence its exclusion from the table. Literature values obtained from plots may be identified by an asterisk.

Figure 9 demonstrates ensemble averaged I-S results for hip joint angles. Angles match well with those reported by other authors [2,28–30,47–49,52,53,90]. This should be expected as the approach to computing joint angles in this study is consistent with that used in other studies. Variability between kinematics computed in this study to those reported in other studies could be attributed to subject-specific biomechanics, differences in gait speed or step height, and differences in biomechanical modeling. The replacement of force
plates with Loadsols in the I-S approach distinguish it from the fully gold standard approach, which would only impact inverse dynamic calculations.

**Figure 9:** Hip joint angles computed using OpenSim Inverse Kinematics for (A) Walking and (B) Stair Ascent. The vertical line at 66% of the gait cycle indicates approximate toe off. The solid line represents the ensemble average and the dotted lines are the average plus and minus the standard deviation. In these plots flexion and adduction are positive whereas extension and abduction are negative.

Figure 10 demonstrates ensemble averaged I-S results for hip joint moments plotted with example curves from prior studies for comparison [28–30]. Moments computed using the I-S approach generally follow the curvature of moments computed using fully gold standard approaches, with a few notable exceptions. In the sagittal plane, I-S computed peak extension and flexion moments are larger and the curve between the peaks smoother than those found in other studies. However, the timing of the peaks within the gait cycle is relatively consistent between I-S and fully gold standard approaches. Peak extension and flexion moments occur at about 20% and 50% of the gait cycle respectively for both activities, except for the peak flexion moment during stair ascent.

In the frontal plane, the I-S approach led to the expected double-peak in abduction during stance. Similar to the sagittal plane moments, the I-S computed peak abduction plane moments were much larger in magnitude than the fully gold standard computed abduction moments. This was consistent across activities. The curve between the double peaks also appears to be flatter for the I-S results compared to the fully gold standard results.
Figure 10: Hip joint moments computed using OpenSim Inverse Dynamics for (A) Walking and (B) Stair Ascent. The vertical line at 66% of the gait cycle indicates approximate toe off. I-S results are presented in blue where the solid line is the ensemble average and the dotted line is the standard deviation. In these plots flexion and adduction are positive whereas extension and abduction are negative. The black lines are examples from studies using a fully gold standard approach for comparison [28–30].

3.2 Stair Ascent on the Exercise Machine

Kinematic and kinetic gait patterns were similar across activities. Notably, in the sagittal plane, stair ascent required a greater range of motion (ROM) than walking (approx. 50° for stair ascent v. 35° for walking), but lower overall moments. Lower magnitude sagittal peak moments in stair ascent may be attributed to the slower gait speed. In the frontal plane, stair ascent required a similar ROM (approximately -5° to 5°) and peak abduction moment during stance to walking. However, stair ascent exhibited a sharper peak adduction angle than walking. For the adduction moment, the first of the two peaks in abduction during stance was slightly larger than the second for stair ascent. The opposite was true for walking.

3.3 Wearable-ANN Approach

3.3.1 ANN Design

The ANN design workflow consisted of three investigations to determine the effects of changing 1) the number of hidden nodes (NHN) in the first hidden layer, 2) the NHN in the second hidden layer, and 3) the number of input nodes. Figure 11 summarizes the results from ANN design investigations. Each bar in these plots represents the average
rRMSE split by output for the ANN of a particular size (i.e. [50,10], [25,10], etc.) across 10 rounds of training, where one round consists of 17 LOO-CV iterations. The rRMSEs split by output across 10 rounds was found to be approximately normally distributed by the Anderson-Darling test in all cases. This allowed for two-tailed t-test comparisons between the rRMSE performance of ANNs of different sizes. Overall, the investigations revealed that the NHNs in either hidden layer was less important than the number of inputs.

**(A) Walking**  
![Graph showing rRMSE split by layer size for Walking](image)

**(B) Stair Ascent**  
![Graph showing rRMSE split by layer size for Stair Ascent](image)
Figure 11: Summary of findings from ANN Design (A) Walking, (B) Stair Ascent; Top: Number of Hidden Nodes (NHN) in Hidden Layer 1 Investigation; * indicates statistically significant difference between adduction moment rRMSE of [25,10] to that of [10,10]. Middle: NHN in Hidden Layer 2 Investigation; * indicates a bar is statistically different from the blue [5,10] bar in same output group. Bottom: Size of Input Layer investigation; * indicates a bar is statistically different from the green [5,5] “all inputs used” bar in same output group.

The first row of plots in Figure 11 shows the effect of changing the NHN in the first layer when the NHN in the second layer is kept at 10. For this investigation, ANNs of one size were compared to the ANN of closest smaller size. On the plot, that means that two-tailed t-tests were conducted to compare [50,10] to [25,10], [25,10] to [10,10], and [10,10] to [5,10], for three total comparisons. To account for errors induced by completing multiple t-tests in sequence, a Bonferroni correction was made to convert the original $\alpha = 0.05$ to $\alpha/m = 0.0167$, where $m = 3$, or the number of total t-tests [91]. With this level of significance, none of the changes in the NHN proved to be significant, except for the improved rRMSE of the adduction moment when [25,10] was compared with [10,10] for walking (Figure 11, Top, asterisk between red and yellow bar in adduction moment group).

While little statistical significance was found between ANN sizes in this investigation, rRMSE seems to trend downward with decreasing NHNs in the first layer. Further, simpler ANNs were much quicker to train (<30 sec to train an individual ANN). The larger number of training iterations completed (10 rounds x 17 LOO-CV iterations) made even small
improvements in individual ANN training time highly convenient. The [5,10] ANN had the lowest rRMSE across outputs and the lowest training computation time. As such it was used in the second investigation for comparison against ANNs with different second layer sizes.

The second investigation consisted of comparing the [5,10] ANN to an ANN with 15 NHN, 5 NHN, and [5,5] NHNs. Results from this investigation are summarized in Figure 11, middle row. An asterisk over a bar indicates that the rRMSE for that ANN output is statistically different than the rRMSE of the corresponding output of the [5,10] ANN with $p \text{ value } < \frac{\alpha}{3} = 0.0167$. For walking training data, removing the second hidden layer (15 or 5 NHN) or reducing the size of the second hidden layer ([5,5] NHN) made no difference. For stair ascent training data, removing the second hidden layer hurt performance whereas reducing the size of the second hidden layer had no statistically significant impact on performance. Consistency between the ANN used for walking data and that used for stair ascent data would make the W-A approach simpler to use across activities. As such, the size of [5,5] NHNs was used for all iterations in the final investigation of the size of the input layer.

The final investigation, considering the size of the input layer, is summarized by the bottom row of plots in Figure 11. Similar to the second investigation, all comparisons are made between an ANN of consistent size, the [5,5] ANN with all input variables, to ANNs with different input layer sizes. The comparisons are [5,5] with all inputs to a [5,5] with the nGRF removed, a [5,5] with the shank IMU data removed, a [5,5] with the thigh IMU data
removed, and a [5,5] with the time (i.e. duration of activity) removed. In total, there are four comparisons, making the level of significance with the Bonferroni correction \( \alpha \frac{n}{4} = 0.0125 \). As could be expected, the nGRF was found to have a significant impact on the ANN’s performance in predicting joint moments in both planes for both activities. Surprisingly, the [5,5] without shank IMU data and the [5,5] without thigh IMU data either outperformed or matched the [5,5] with all inputs for both walking and stair ascent. Excluding the thigh IMU data (yellow bar in Figure 11, bottom plot) reduced the rRMSE for both the walking flexion and adduction moments, but did not change performance for stair ascent. Removing the shank IMU data (red bar in Figure 11, bottom plot) improved rRMSE for the walking adduction angle and moment, but hurt performance for stair ascent flexion angle. Excluding time, or the duration of an activity, as an input made no statistically significant difference in rRMSE results. Because thigh IMU data improved ANN performance for walking training data, it was removed from the final ANN iteration. For stair ascent, thigh IMU data was kept as an input because excluding it decreased the R² across outputs.

The investigations concerning ANN size and performance led to a final ANN architecture of [5,5] that uses all inputs except for the thigh IMU data for walking and a [5,5] that uses all inputs for stair ascent (See Appendix A for diagram of final ANN architecture). The ANNs were found to perform well for both walking and stair ascent data and could be trained quickly (<10 sec/individual ANN trained, <3min/round of 17 LOO-CV iterations).
3.3.2 ANN Performance

Figure 12 shows that the ensemble averaged W-A results follow the curvature of the I-S results very well. The greatest deviations between results from the two approaches are found at the peaks, where the W-A results consistently undershoot. This undershooting can be found in the results from other studies seeking to use IMUs and ANNs to compute joint kinematics and kinetics [40,43].
Figure 12: Ensemble averaged results for (A) walking and (B) stair ascent hip joint angles and moments in the sagittal (Flexion +, Extension -) and frontal planes (Adduction +, Abduction -). I-S results are shown in blue and W-A results in red. The solid lines represent the average ensembled result and the dotted show the average plus and minus the standard deviation.

To compute final performance metrics, the final ANN architectures were trained for 10 rounds, where each round represents 17 LOO-CV iterations. The metrics, namely R² and rRMSE, were averaged across LOO-CV iterations for each of the 10 rounds, and then averaged again across rounds. The standard deviation was also computed across rounds. Table 3 summarises these metrics. The rRMSEs for all outputs except adduction angle are below the goal 20%. The adduction moment prediction was the most successful, with an R² > 0.92 and rRMSE < 13% for both activities. The adduction angle for stair ascent was the most challenging to predict, resulting in an R² = 0.59 and rRMSE =26.36%. The standard deviation across ensemble averaged adduction angle results for the W-A approach is distinctly narrower than that for the I-S results.

<table>
<thead>
<tr>
<th></th>
<th>Walking</th>
<th>Stair Ascent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rRMSE (%)</td>
<td>R²</td>
</tr>
<tr>
<td>Flexion Angle</td>
<td>16.04 ± 0.5</td>
<td>0.85 ± 0.01</td>
</tr>
<tr>
<td>Adduction Angle</td>
<td>20.76 ± 0.47</td>
<td>0.79 ± 0.02</td>
</tr>
<tr>
<td>Flexion Moment</td>
<td>11.24 ± 0.69</td>
<td>0.83 ± 0.02</td>
</tr>
<tr>
<td>Adduction Moment</td>
<td>11.65 ± 0.61</td>
<td>0.93 ± 0.01</td>
</tr>
</tbody>
</table>

Table 3: Tabulated performance metrics for W-A results reported as the average ± SD across rounds.

This study follows Mundt et al. in deciding not to remove statistical outliers. However, subjects with average rRMSE across the 10 rounds that fell 1.5 interquartile ranges above the 75th percentile or below the 25th percentile were investigated by plotting ensemble averaged, I-S computed hip joint angles and moments by subject (Appendix B). The subjects with particularly high rRMSEs were highlighted by plotting with bold, colored lines. Their I-S results did not differ ostensibly from the rest of the subjects. As such, removing outliers would have likely removed natural biomechanical variability and might
misrepresent the capability of the trained ANNs to predict accurate hip joint kinematics and kinetics on new subjects. Violin plots in Figure 13 demonstrate variability in the W-A performance across subjects. Each dot in the plot represents the average performance of that subject as the validation set across rounds. The white dot represents the median value and the grey bars the interquartile range. The violin shape is constructed by calculating the kernel density estimation of the probability distribution curve of the dataset. The W-A approach predicted sagittal plane outputs well for most subjects (e.g. fat bulb near $R^2=0.90$ for walking flexion angle and moment), and poorly for a handful of others. Adduction moments were predicted with high accuracy across subjects and activities. The accuracy of predicting adduction angles varied more widely across subjects.

*Figure 13: Violin plots of performance metrics for W-A method. Each dot represents the average output metric for a particular subject acting as the validation set across 10 rounds of ANN training. The white dot at the center of the violin is the median value, and the grey bar shows the interquartile range. The width of the violin is determined by fitting a kernel density estimation to show the probability distribution curve.*

4. Discussion

The goal of this study was to develop two new methods for computing hip joint angles and moments in the sagittal and frontal planes. The proposed methods improve the cost, portability, and convenience of hip joint kinematic and kinetic analysis. MOCAP and force
plate prices range in the tens of thousands whereas commercially available wearables typically cost in the thousands range [25–27]. Further, some researchers have developed simple wearables for even lower cost (<$1000 [92,93]). The Insole-Standard (IS) approach leveraged force-measuring, ‘Loadsol’, insoles in the place of force plates to allow for a ‘pseudo’ gold standard approach. These insoles are more convenient to use than force plates. To measure GRFs during gait, walkways or steps are typically constructed with force plates embedded. This is costly, inconvenient, and commonly allows for the capture of only a single gait cycle at a time. In contrast, the Loadsols are slipped into subjects’ shoes, allowing for data capture on exercise machines and the capture of many gait cycles to be captured in succession, better emulating human motion outside of the laboratory. The Wearable-ANN (W-A) approach represents a fully portable alternative to the conventional workflow used to compute hip joint kinematics and kinetics. Using just one to two IMUs, the Loadsols, and a shallow, [5,5] ANN, the W-A approach produced results that matched the curvature well for most outputs (Average $R^2 = 0.77$ across outputs and rounds).

4.1 Insole-Standard Approach

To the author’s knowledge, no study has reported using Loadsols with one sensing zone measuring one total nGRF per sensor and OpenSim to compute hip joint moments. Considering the limitations of the Loadsols, the curvature of the I-S computed results matches fairly well with results from studies using a fully gold standard approach. However, peak flexion moments were more than 25% greater than those found in by Costigan et al. [29], disproving one hypothesis in this study. Further, I-S computed peak moments were consistently much larger than those found in the literature. This suggests
that center of pressure (CoP) and/or non-vertical GRF (i.e. medio-lateral, anterior-posterior forces) information are critical to computing joint moments with the correct peak magnitudes. The I-S results in this study are informative for analyzing the effects of completing ID computations without full GRF information.

First, peak moment magnitudes in the I-S results likely show the effect of holding the application point of the nGRF at the talus rather than modeling the trajectory of the CoP during gait. While the magnitude of the nGRF captured by the Loadsols is consistent with that captured by force plates in other studies for the same activities (Appendix C; [28,49,90]), applying it to the talus changes its lever arm about the hip joint center. McCaw et al. analyzed the sensitivity of ID to small errors in CoP [46]. They found that anteriorly shifting the CoP one centimeter from where it should be caused an 8% increase in the maximum extension moment computed using inverse dynamics. Shifting the CoP one centimeter posterior to where it should be caused a 16% increase in the maximum flexion moment. Holding the nGRF application point at the talus during gait would represent an anterior shift from the real CoP of a couple centimeters during initial stance phase and a posterior shift of several centimeters during terminal stance phase (Figure 14). Chiu et al. reported an average foot length of 24.5 ± 1.4 cm and foot width of 8.9 ± 0.6 cm across 30 subjects (15M) and a CoP trajectory during gait that took up 95% of the foot length and 31% of foot width ([94]; Figure 14). The time of peak flexion moment during terminal stance would likely represent a posterior shift of 15 cm or more. If every centimeter difference between the true CoP and the talus represented a 16% difference in peak flexion moment, like in the McCaw study, an error of 15 cm would account for a 240% increase.
The peak flexion moments computed using the I-S approach were two to four times greater than those computed using fully gold standard approaches. A 240% increase could account for the differences in peak flexion moment.

![Figure 14: (A) Figure to help explain error induced by applying the nGRF to the talus. Solid arrows represent the nGRF (red = this study, blue = nGRF at true CoP). Dotted lines show the line of action and lever arm of the nGRF about the center of the hip. Gait figure adapted from Tekscan Inc. [95]. (B) The trajectory of the CoP during stance from Chiu et al. [94]](image)

Similarly, at the time of peak extension moment, just after heel strike, the true CoP would likely be within a few centimeters of the talus, no more than 10. If each centimeter between the real CoP and the talus accounted for an 8% change in the peak extension moment during initial stance, a 10 cm difference could account for an 80% increase in peak moment. Most authors report a higher peak extension moment than peak flexion moment during walking and stair ascent [30,47,52]. The I-S approach computed peak flexion and extension moments of similar magnitude. It is possible that error in CoP position caused a greater increase in peak flexion moment than in peak extension moment, making them appear to be roughly equal in magnitude.

Kim et al. confirmed the findings by McCaw et al. about shifting the CoP in the sagittal plane and extended the analysis to understand CoP shifts in the frontal plane [96]. Kim found that shifting the CoP laterally increased the double-peak abduction moment found
during stance phase for walking. In his study, a 3 cm lateral shift corresponded with an approximately 20% higher double-peak in abduction. Holding the CoP at the talus would likely represent a lateral shift in the CoP for most of stance phase, which may explain why the I-S approach resulted in higher-than-expected abduction moments.

Anteroposterior and mediolateral GRF components represent a small fraction of the total GRF for both walking and stair ascent [49]. It is likely that their small magnitudes make it difficult to conduct a sensitivity analysis on ID results like that conducted by McCaw et al. and Kim et al. with CoP errors. Most papers studying normal, healthy subjects do not mention the impact of non-vertical GRFs on ID results. However, studies of subjects with gait pathologies have found non-vertical GRFs to be useful metrics to compare against healthy subjects [97,98]. Therefore, it is likely that the I-S approach would need to accurately account for nonvertical GRFs if it were to be used to study subjects with gait pathologies.

While I-S results exhibited higher peak moments, the curvature matched that of results from similar studies using a fully gold standard approach. Future studies are needed to compare the I-S approach to a fully gold standard approach and to address the limitations imposed by the Loadsols. There are a few simple options to addressing these limitations. To start, the CoP trajectory could be modeled off of prior work, like that of Chiu et al., for walking [94]. Chiu normalized the CoP to foot length and width, tracking the position over percent time in the gait cycle. The absolute position on a subject’s foot could be found by multiplying Chiu’s normalized CoP position by their foot length and width. Once time is
converted into percent of the gait cycle, it would be easy to match this approximate CoP trajectory to the nGRF. Custom MATLAB scripts, like those developed in this study, could be used to write this approximate CoP into .mot files for import into OpenSim. Alternatively, Hullfish et al. proposed another CoP estimation for their study which used Loadsols with three sensing zones, each measuring the a part of the nGRF, to compute ankle moments using OpenSim [99]. This study picked a position within each of the three Loadsol sensing zones, calculated the moment arm from the ankle joint center, and subtracted a constant offset, the distance between the ankle joint center and the posterior aspect of the Loadsol. With this CoP estimation, they achieved accurate ankle joint moment results compared to a fully gold standard approach ($R^2 = 0.98$).

Finally, one other approach could be used to address both the lack of CoP and nonvertical GRF information: replace the force-measuring insoles with pressure-measuring insoles. Many biomechanists have begun to develop algorithms for calculating the CoP and full three-dimensional GRFs from plantar-pressure insoles [44,100–102]. While these sensors are typically more expensive, they offer increased information over the Loadsol force-measuring insoles with the same portability.

### 4.2 Stair Ascent on Exercise Machine

Greater continuity between gait cycles likely characterizes out-of-laboratory stair ascent better than a 3-5 step system. Capturing stair ascent on the stair exercise machine was also more convenient. It did not require the construction of specialized equipment (i.e. steps with force plates embedded) and took a short period of time to capture many gait cycles.
However, the biomechanics of stair ascent on the stair exercise machine are different than those for stair ascent on a normal set of steps. Two characteristics of the stair master contribute to these differences.

First, the machine’s steps are higher and narrower than that of a standard set of stairs (machine: 8 in. rise x 9.25 in. run; standard: 7 in. rise x 11 in. run), and second, the steps move away from the posterior side of the subject as they ascend. The above-standard step height corresponds with a higher stair incline, which has been shown to increase the hip extensor moment and the hip flexion angle during ascent compared to that completed on a set of stairs with a lower incline [49]. This increase in hip extension moment due to the high incline of the stair case may be counteracted by the movement of the steps. On a normal, nonmoving set of steps, the hip generates an extension moment during stance to accelerate the center of mass forward. On the stair exercise machine, the subjects’ center of mass oscillates rather than primarily moving forward. After heel strike, the leg in contact with the ground is accelerated backwards by the machine, likely alleviating the extension moment typically generated during early stance phase. This backward acceleration likely has the opposite effect on flexion moment. Because the step is moving posteriorly, the subject must generate an extra large flexion moment to lift their leg up from one step to clear the next one.

Differences between the kinetic results of stair ascent and walking in this study are confounded by error in keeping the nGRF application point at the talus. Future work would need to estimate the CoP for stair ascent like Chiu et al. did for normal walking. Lee et al.
characterized the CoP-Center of Mass inclination angles during stair ascent, but these parameters may not represent CoP trajectory on the stair machine [103]. If the stair exercise machine were to be used in future studies, it is likely a pressure-measuring insole would be required to accurately characterize CoP trajectories. This is particularly important for subjects with long foot lengths. The steps on the stair exercise machine did not provide enough depth for all subjects to place their feet flat on the step. Subjects with particularly long feet (> 30 cm) were forced to tip-toe as they ascended the stair case. This would have severely limited the CoP trajectory in these subjects compared to subjects with shorter feet.

4.3 Wearable-ANN Approach

4.3.1 ANN Design

The study found that a shallow ANN of size [5,5] worked best for predicting hip kinematics and kinetics for both activities. The only difference in architecture between the final iteration of the ANN for predicting walking results from that predicting stair ascent results was the exclusion of thigh IMU data as inputs from the walking ANN. The size of the training dataset, the number of inputs, the number of outputs, and the complexity of the relationship between inputs and outputs all contribute to determining the optimal number of hidden nodes (NHN) in an ANN [88]. In this study, a small training dataset and few input variables likely contributed to the success of the smaller ANN over one with a greater number of hidden nodes (NHN). Hecht-Nielson suggested that the optimal NHN be guided by the following relation: \( NHN \leq N_{INPUT} + 1 \). Similarly, Masters et al. suggested ANN architectures should resemble a pyramid, with NHN approximately equal to \( \sqrt{N_{INPUT} \times N_{OUTPUT}} \). The final ANNs in this study fit these guidelines quite well. Both the
walking and stair ascent ANNs had 10 NHN, which is less than the input nodes used plus one for both activities. Regarding Master’s proposed guideline, the value of \( \sqrt{N_{\text{INPUT}} \times N_{\text{OUTPUT}}} \) would be 8 for walking and 13 for stair ascent. This is close to the 10 NHN used. Further, the final ANN iterations in this study resemble a pyramid shape, with the greatest nodes in the input layer, fewer in the hidden layers, and fewer still in the output layers.

ANN design for the estimation of joint kinematics and kinetics can be guided by biomechanical modeling, like in the study completed by Lim et al., but it is typically not. Perhaps the size and variability of the training dataset and the number of input and output nodes is more clearly linked to the ideal NHN for an application than the standard state space equations used to model dynamic biomechanical motion. Lim et al. used a linked segment model to choose seven input variables from data captured by an IMU on the waist to predict eleven kinematic and kinetic output variables. The linked segment model may have proved useful for choosing the input nodes, but ultimately, Lim et al. used trial and error to determine the NHNs. Further, their study was constrained to analyzing sagittal plane motion. Opening up the analysis to motion in other planes makes biomechanical modeling more complex, and harder to relate to ANN computations. Attempting to reconcile ANN computations to biomechanical modeling may not be as useful an endeavor as understanding the statistical patterns in training data ‘learned’ by ANNs and how that relates to the ANN’s ability to then predict outputs on new data.
Mundt et al. found that soft tissue noise in IMU data improved the prediction accuracy of their ANN computing lower body joint moments in all three planes [39]. Further, they found that the amount of soft tissue noise in IMU data seemed to have a greater impact on improving ANN performance than increasing the size of the training dataset. Mundt does not attempt to explain this finding, but it could be postulated that greater soft tissue noise in a subjects’ data may be indicative of a higher mass, which may be used by ANN to better predict the magnitude of joint moments. Interestingly, although ANNs are known to require large training datasets, the results in this study and in the study by Lim et al. seem to indicate that shallow ANNs may be trained with relatively small datasets (<100 gait cycles for stair ascent in this study) to achieve success ($R^2 > 0.70$ for joint moments in both the sagittal and frontal planes). The success of smaller datasets may be attributed to including enough variability in the training dataset to well encompass the patterns in the test or validation datasets. In the study by Lim et al., subjects were similar in terms of age, height and mass (SD of 2.9 years, 7.7 kg, and 7.5 cm respectively). The subjects were instructed to walk at three gait speeds, which were then combined for LOO-CV training. The similar age and mass characteristics among subjects combined with variable gait speeds may have allowed for relatively accurate prediction. In this study, subjects walked and ascended stairs at a controlled speed. Variability in the training data may be attributed to larger standard deviations in age and mass across subjects (SD of 6.35 years and 19.5 kg in this study compared to 2.9 years and 7.7 kg in the study by Lim et al.). The combination of controlled gait speed with variable subjects seem to provide successful ANN performance like that in the study by Lim et al.
4.3.2 ANN Performance

Across outputs and performance metrics, the W-A results perform well against other studies using IMUs and ANNs to compute joint moments and angles. It achieves an $R^2 > 0.75$ and an rRMSE $< 18\%$ for all outputs but the adduction angle. Mundt et al. achieves an $R^2 > 0.90$ and an rRMSE $< 15\%$ for all outputs from their ANN, but uses five IMUs, thousands of hidden nodes, and thousands of training iterations. Lim et al. achieves an $R^2 = 0.81$ and rRMSE $= 10.74\%$, but uses a larger data set and only computes hip joint moments in the sagittal plane. Few studies combine instrumented insoles, force- or pressure-measuring, with ANNs for predicting joint moments, and even fewer combine instrumented insoles, inertial measurement units, and ANNs. The W-A approach in this study leveraged Loadsol force-measuring insoles, which proved valuable for predicting moment outputs (Figure 11, bottom plots; Section 3.3.1 ANN Design).

Pressure-sensing insoles may offer even greater advantages to developing wearable-ANN approaches. Jacobs et al. found success (rRMSE $< 10\%$ for all outputs) using a custom developed pressure-sensing insole for data capture and an ANN for the prediction of anterior-posterior GRFs, CoP positions, and ankle joint moments [93]. His study did not use IMUs. Future work should investigate the advantages of using different wearables (i.e. IMUs, force-measuring insoles, pressure-sensing insoles) for ANN-based joint kinematic and kinetic prediction. Excluding either shank or thigh IMU data improved performance in the walking ANN, but excluding shank data hurt performance in the stair ascent ANN. As such, the ideal quantity, type, and location of wearables for lower body joint angle and moment prediction may depend on the activity being studied.
Wearable-ANN methods used to predict hip joint angles and moments thus far have found that ANN-estimated outputs exhibit lower variance and peak magnitudes than outputs computed using conventional biomechanical methods (i.e. ‘ground truth’, IK/ID results). In this study, differences in peak magnitudes between the I-S and W-A approaches are most evident in flexion moments for both activities, where peaks are most narrow. Difference in variance between the I-S and W-A approach are most apparent for the adduction angle, the poorest performing output for both activities, but especially for stair ascent ($R^2=0.59$). Subjects appear to exhibit higher kinematic variation in the frontal plane, making it more difficult to achieve high accuracy using the W-A approach. The ANN seems to have difficulty reconciling these widely varying kinematics. Interestingly, W-A computed adduction moments achieve the highest performance across outputs ($R^2>0.90$). Net hip moments would be more important to THA failure analyses, but the inability of the W-A approach to achieve high accuracy with more variable output data would represent a weakness of the approach that could compromise its usefulness in other applications. The author agrees with Mundt’s suggestion that future efforts in developing ANNs for the prediction of lower body joint angles and moments should consider data augmentation to address this challenge [39]. Data augmentation is commonly used in deep learning to classify images and consists of transforming training images by rotating, cropping, and zooming. The analog to data augmentation in image classification for IMU data could consist of randomly rotating the relative orientation of the sensor as Mundt et al. did to their ‘simulated’ IMU data in their 2019 paper [39].
ANN performance could also be improved by designing for a specific subject set. Subject-specific ANNs perform better than ANNs trained using one subject set on another subject set [104], but are less convenient because they require collecting new training sets for each subject. That would require completing the gold standard approach for each new subject to establish the ground truth, which would eliminate the advantage of portability and lower costs. Instead, ANNs could be designed to predict joint kinematic or kinetic outputs for a particular subject class, split by gender, age, or pathology for example.

Lastly, the sensitivity of ANN performance to syncing errors has yet to be characterized. This study uses gait analysis to manually identify the first heel strike in each type of data (i.e. MOCAP, Loadsol, IMU). Preliminarily, two subject’s IMU data failed to sync well with data from the other two sensing modalities, resulting in extremely poor performance when included in data analysis. This syncing error was fixed and their performance improved dramatically, improving overall final results. Other researchers use alternative syncing approaches. Mundt et al. uses an algorithm to minimize the mean-square objective function of the medial-lateral acceleration of the pelvis between MOCAP and IMU data [40]. Jacobs et al. recorded the manually triggered square wave from a signal generator on each data acquisition system and used that for syncing [93]. Neither Mundt nor Jacobs note the effectiveness of their syncing protocol or its impact on the performance on their ANN’s performance. Characterizing the sensitivity of ANN performance to syncing effectiveness and using consistent syncing methods across studies would improve the development of wearable-ANN methods.
The W-A approach proposed in this study demonstrated the capability of a simple ANN with small training data sets for calculating joint angles and moments that match those computed using biomechanical modeling tools. Future work is needed to compare the W-A approach to a fully gold standard approach. Even so, the success of the W-A approach in matching the curvature of the I-S approach is encouraging. There are many opportunities to consider improving wearable-ANN approaches to estimating joint kinematics and kinetics, including determining the ideal quantity, type, and location of wearable sensors for the particular biomechanical output of interest, implementing data augmentation techniques, refining the definition of the target group (i.e. male/female, young/older, healthy/has osteoarthritis), and characterizing the sensitivity of ANN performance to the effectiveness in sensor synchronization.

4.4 Significance

Ultimately, the goal of this study was to develop a fully portable, more convenient method for computing hip joint angles and moments using wearables and ANNs. The study was relatively successful in establishing the correct curvature in both kinematic and kinetic outputs using a ‘pseudo’ gold standard approach, the I-S method. Next steps were suggested to improve the I-S approach and achieve correct kinetic peak magnitudes.

The proposed W-A approach required data capture from only 1-2 IMUs and one force-measuring insole. This represented an improvement over attempts to compute lower body joint angles and moments using IMUs and traditional biomechanical methods, which require many IMUs (>4) and complex coordinate transformations and achieve relatively
modest results ($R^2 < 0.60$, rRMSE > 25%) [35–37]. The optimization algorithms used by studies combining IMUs with traditional biomechanical modeling tools also require long computation times. In contrast, once ANNs are trained, new predictions can be made within seconds, making real-time computations possible and long duration (i.e. hours or days of data) analysis more convenient.

These quick computation times extend the applicability of the W-A method beyond THA failure analysis. For example, W-A computed kinematics and kinetics could potentially be used for gait retraining [47,106]. Gait retraining requires measuring real-time kinematic or kinetic data during an activity to determine what cues (i.e. visual, auditory) should be given to the patient to help them modify their movement and to assess the effectiveness of those cues in causing positive change. To be useful for gait retraining, the W-A approach would need to be extended to compute additional joint angles and moments beyond those around the hip. This would not require complex method development. The same ANN design workflow described in this study to design and ANN for computing hip joint metrics could be used to design an ANN to compute other joint metrics. With this improvement, the W-A approach could be used for gait retraining of THA patients with Trendelenburg gait, for example. Trendelenburg gait is characterized by drooping of the pelvis on the unloaded side during walking due to weak abductors. Hamacher et al. showed that a visual feedback system using kinematic data improves Trendelenburg gait in THA patients [107]. Another application for using the W-A approach and gait retraining includes teaching total knee arthroplasty (TKA) patients to increase their mediolateral trunk sway during walking, which has been shown to reduce the knee adduction moment. Reducing this moment can
decrease the patient’s risk of developing medial compartment knee osteoarthritis [108]. Trendelenburg gait in THA patients and medial compartment knee osteoarthritis in TKA patients represent two of many pathologies where the W-A approach could be leveraged to improve gait retraining, making treatment outside of the clinic or laboratory possible.

The W-A approach offers reduced cost, greater portability, and faster computation times over a fully gold standard approach to computing hip kinematics and kinetics. This study highlighted two applications that would leverage W-A advantages for treating arthroplasty patients, THA failure analysis and gait retraining. The advantages of the approach make it applicable elsewhere, for rehabilitation of patients with other pathologies causing abnormal gait for instance. Ultimately, with further work to validate the approach against a fully gold standard approach and extend it to predict joint angles and moments more broadly, it has the potential to impact millions of arthroplasty patients in the United States, as well as many other patients suffering from gait pathologies.
Appendices

Appendix A: Diagram of Final ANN Architectures

* Walking ANN did not use Thigh IMU input nodes
Appendix B: Ensemble averaged I-S results across subjects

Figure B1: I-S results ensemble averaged and split by subject. Outliers were determined as falling greater than 1.5 interquartile ranges away from the 75\textsuperscript{th} or 25\textsuperscript{th} percentiles. They are indicated by colored, bolded lines (Cyan = Subject 11, Magenta = Subject 17)

Appendix C: Ensemble averaged nGRFs across Subjects

Figure C1: Ensemble averaged nGRF for (A) Walking and (B) Stair Ascent.
References


[45] Fong, D. T. P., Hong, Y., Yung, P. S. H., Fung, K., Lao, L. M., Chan, K., Kong,


University, S., “OpenSim Documentation: How Scaling Works.”

University, S., “OpenSim Documentation: How Inverse Kinematics Works.”


