ASME Journal of Biomechanical Engineering
Online journal at:
https://asmedigitalcollection.asme.org/biomechanical

Mirel Ajdaroski
Department of Mechanical Engineering,
University of Michigan – Dearborn,
4901 Evergreen Road,
Dearborn, MI 48128
e-mail: majdaros@umich.edu

So Young Baek
Department of Mechanical Engineering,
University of Michigan,
Ann Arbor, MI 48109
e-mail: soy@umich.edu

James A. Ashton-Miller
Department of Mechanical Engineering,
University of Michigan,
Ann Arbor, MI 48109
e-mail: jaam@umich.edu

Amanda O. Esquivel
Department of Mechanical Engineering,
University of Michigan - Dearborn,
4901 Evergreen Road,
Dearborn, MI 48128
e-mail: aoe@umich.edu

Predicting Leg Forces and Knee Moments Using Inertial Measurement Units: An In Vitro Study

We compared the ability of seven machine learning algorithms to use wearable inertial measurement unit (IMU) data to identify the severe knee loading cycles known to induce microdamage associated with anterior cruciate ligament rupture. Sixteen cadaveric knee specimens, dissected free of skin and muscle, were mounted in a rig simulating standardized jump landings. One IMU was located above and the other below the knee, the applied three-dimensional action and reaction loads were measured via six-axis load cells, and the three-dimensional knee kinematics were also recorded by a laboratory motion capture system. Machine learning algorithms were used to predict the knee moments and the tibial and femur vertical forces; 13 knees were utilized for training each model, while three were used for testing its accuracy (i.e., normalized root-mean-square error) and reliability (Bland–Altman limits of agreement). The results showed the models predicted force and knee moment values with acceptable levels of error and, although several models exhibited some form of bias, acceptable reliability. Further research will be needed to determine whether these types of models can be modified to attenuate the inevitable in vivo soft tissue motion artifact associated with highly dynamic activities like jump landings. [DOI: 10.1115/1.4064145]

1 Introduction

In the U.S., approximately 100,000–200,000 anterior cruciate ligament (ACL) ruptures are reported every year [1,2]. Several studies have examined the mechanisms by which an ACL may be injured, usually during sports like basketball, football, soccer, and volleyball, including so-called "contact" (30% of all cases) and "noncontact" (70% of all cases) injuries [3–6]. Noncontact injuries involve no load transfer from another player but rather loads on the knee generated by limb muscles and passive structures as well as the ground reaction force beneath the foot. Recent studies have suggested that some noncontact ACL injuries may be due to overuse of the ligament, whereby repetitive high-stress/strain-inducing activities result in ACL collagen unraveling and multiscale damage thereby weakening the ligament; without adequate built-in rest periods for repair, ACL rupture can occur [7–10]. So, counting the accumulated number of these potentially injurious loading cycles over a given period could have considerable clinical import, as we shall discuss later.

Older studies have reported the circumstances of noncontact ACL injury. These include landing with an externally rotated knee in an extended position (20 deg or less of flexion) in slight valgus, or with a powerful quadriceps contraction; this can cause anterior displacement of the tibia with respect to the femur via the patella-femoral mechanism, thereby increasing the ACL strain above its tolerable threshold [11,12]. Other studies suggest that landing a jump with a fully extended knee in neutral, or internal rotation with valgus orientation, can excessively strain the ligament [11], with ACL stress being correlated with the magnitude of internal rotation and valgus knee moments [13–16].

Traditionally, lower limb joint kinematics and kinetics are measured in laboratory settings using camera-based motion capture systems and force plates. However, such systems are large and bulky, making them impractical for regular usage in the field. In recent years the development of wearable sensors has emerged as an alternative to camera-based motion capture systems, and they have usefully been employed in various applications including gait analysis and assistive rehabilitation [17–23]. Inertial measurement unit (IMUs) are small, wearable sensors generally consist of orthogonal accelerometers, rate gyroscopes, and magnetometers allowing direct measurement of a body segment's linear acceleration, angular velocity, and environmental magnetic field strength. IMUs, of course, can be used as a system to capture information about multiple body segments to provide information about physiological angles of various joints. Studies have also examined the correlation between linear acceleration and ground reaction force (GRF), joint force, and moments in the lower limb with varying degrees of agreement [24–29]. Various techniques ranging from regression modeling to developing machine learning algorithms have been applied [30]. To develop models that can be used to track events on-field, one might focus on factors that can be obtained from IMU data, and perhaps combine them with variables measured before testing (i.e., weight, height, etc.). GRF and knee moments are of interest because many actions that can cause noncontact ACL injuries are those in which a change of direction or sudden deceleration occurs thereby applying significant orthogonal moments to the knee in the sagittal, coronal, and/or transverse

Manuscript received March 17, 2023; final manuscript received November 14, 2023; published online December 20, 2023. Assoc. Editor: Bruce MacWilliams.

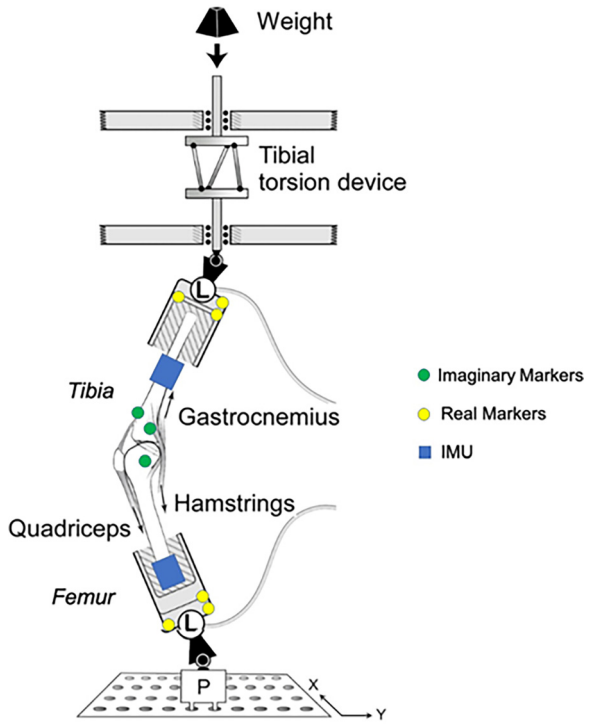

Fig. 1 Schematic of the test-rig. IMU placement and digital marker location labeled. Reproduced with permission from Ref. [34].

planes; forces in such cases may be less important than certain combinations of moments experienced by the knee [31–33].

The purpose of this study was to develop multivariate models using data that are available from two wearable IMU sensors, one located above and the other below the knee, in order to predict GRF and knee moments in cadaveric specimens. The rationale for using whole knees in vitro was twofold. First, this allows us to apply loading cycles that are known to cause fatigue damage to the ACL to examine how IMUs can be used to identify these potentially injurious loading cycles. This method also allows us to focus on developing models to estimate impulsive forces and moments using IMUs under ideal conditions—without the inevitable soft tissue movement artifact present in vivo. This advance, in turn, could encourage athletes, coaches, and trainers to make changes in athletic training loads in order to permit adequate rest and recovery; they could also adjust a training regimen to prevent an ACL rupture from occurring due to possible overuse.

2 Methods

2.1 Specimens, Instrumentation, and Testing Procedures. Sixteen cadaveric knee specimens were used to develop and test each constructed multivariate model (Table 1). Each specimen was cut to a standard length of 20 cm, dissected of tissue while leaving the ligamentous capsule and key tendons intact, and then potted in polymethylmethacrylate cylinders. Prepped specimens were mounted in a custom-built and validated testing apparatus designed to simulate a one-legged landing in the presence of simulated trans-knee muscle forces (Fig. 1) [34,35]. Specimens were instrumented with two wearable IMUs (APDM Opal, APDM Wearable Technologies, Portland, OR), one rigidly attached to the medial aspect of the midtibia and the other to the lateral aspect of the midfemur using a combination of Coban™ and elastic ties. Two six-axis load cells (MC3A-1000, AMTI, Watertown, MA) were used to measure action and reaction forces and moments applied to the knee. A Certus optoelectronic tracking system (Optotrak Certus; Northern Digital, Inc., Waterloo, ON, Canada), which served as the motion capture system, was used to measure knee kinematics and to determine the moment arms for individual knee specimens about each orthogonal axis. For the flexion moment, the moment arm was the distance from the lateral epicondyle to the tibial load cell; for the abduction moment it was the distance between the medial and lateral epicondyles, while for the rotation moment it was directly measured by the tibial load cell. All models that we developed were based on the data of the certus/load cell system (CLS). The CLS sampling rate was 2 kHz while that of the IMUs was 200 Hz.

After each specimen was positioned, the IMU sensors were calibrated and initialized through the predefined calibration conditions used by the MOVEO MOBILITY software developed by APDM, after which, the sensors continuously recorded data until testing concluded. The quadriceps, hamstring, and gastrocnemius muscles were pretensioned to 180 N, 70 N, and 70 N, respectively (mimicking real-world values), the initial knee flexion angle adjusted to 15 degrees, and the quadriceps tendon was clamped securely in place with mechanical grips cooled by liquid nitrogen. This last step was repeated every 25 cycles to prevent tendon thawing and slippage from the clamp; it was only during these instances that corrections of the knee angle and/or muscle tension occurred.

Landing force magnitude was fine-tuned by adjusting drop height/weight to achieve the desired submaximal impulsive knee loading, ranging from one to four body weights (1x–4xBW) applied to the tibia. A combination of paired and single-knee specimens was used. Paired knee specimens had one side (either left or right) tested under lower BW conditions (1–2xBW) and the other at higher (3x+ BW), whereas single knee specimens were all tested under high submaximal loading conditions (3x–4x BW). Impacts were initiated by releasing a weight, guided via two parallel linear bearings, to impact the load cell at the distal end of the tibia (Fig. 1). Based on previous studies, at least five preconditioning trials were conducted before the activation of the tibial torsional device. This allowed for any test rig adjustments and for potential uncrimping of ligament collagen fibers [34,35].

After the preconditioning trials the tibial torsional device was then activated allowing for some of the impact force to be converted

Table 1 Specimen demographics

Specimen ID	Sex	Age	Mass (kg)	Total testing trials	Trials analyzed
Demographics of the specimens used in model training					
P1	M	33	72.6	216	200
P2	M	25	68.9	118	200
P3[a]	M	32	64.4	136	119
P4	M	19	71.7	226	200
P5	M	20	69.9	214	200
P6	M	25	80.3	221	100
S1	M	39	60.8	98	88
S2	M	37	77.1	110	100
S3	M	38	69.9	126	96
S4	M	33	80.3	120	100
S5	M	15	68.9	105	99
S6	F	25	86.2	105	100
S7	F	39	54.4	105	100
Demographics of the specimens used in model testing					
T1	M	23	72.6	105	100
T2	M	23	72.6	108	100
T3	F	26	68.9	105	96

[a]Specimen experienced a failure before the completion of 100-cycle testing due to aTT > 3 mm.

A total number of testing trials denotes how many trials the specimen was able to undergo before failure or completion of 100-cycle testing occurred. Trials analyzed are the number of trials where proper testing was performed; these are the trials that were used in model development/testing. Specimens denoted with "P" are paired specimen; total testing trials and trials analyzed are the sum of the pair's testing sessions.

Table 2 Complete overview of the best-performing models for each of the seven metrics

Model	Type	Kernel function	Layers (size)	Basis function	Activation
		Machine learning models			
Vertical tibial force	GPR	Isotropic rational quadratic	–	Constant	–
Vertical femoral force	Narrow neural network	–	1 (10)	–	ReLU
Resultant tibia force	GPR	Isotropic exponential	–	Zero	–
Resultant femoral force	Bi-layered neural network	–	2 (10)	–	ReLU
Flexion moment	GPR	Nonisotropic 3/2 matern	–	Zero	–
Abduction moment	GPR	Nonisotropic rational quadratic	–	Constant	–
Rotation moment	SVM	Gaussian	–	–	–

GPR model type is a Gaussian process regression model, while SVM indicates a support vector machine learning algorithm. The activation type ReLU is the rectified linear unit.

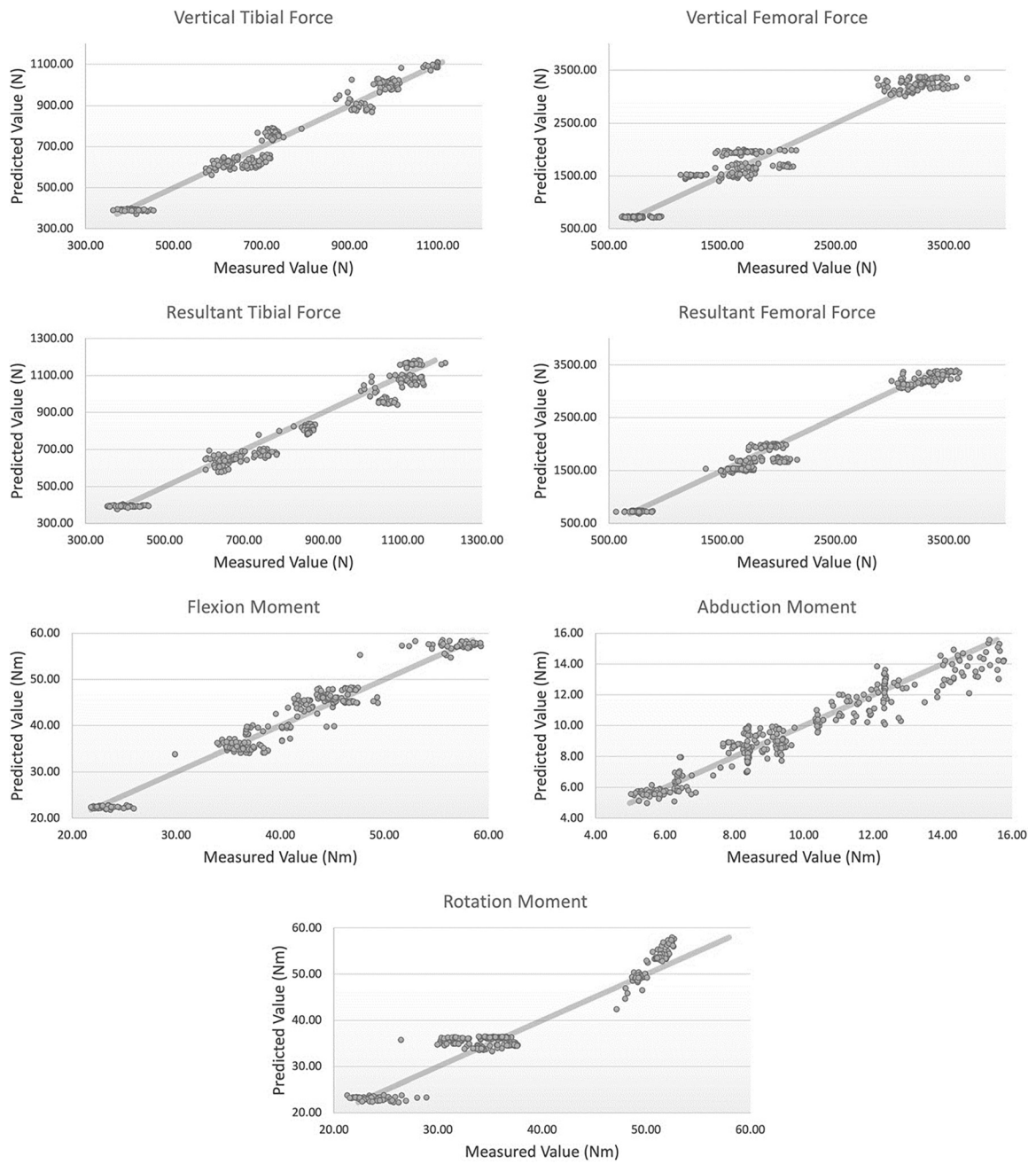

Fig. 2 Relationship between the predicted and measured value for each of the seven parameters

to an internal tibial torque, as can happen in a live subject when the GRF causes internal tibial rotation at the knee [34,35] (Fig. 1). Each specimen was then subjected to up to 100 simulated landing trials or until knee failure occurred, the latter clinically defined as anterior tibial translation (aTT) exceeding 3 mm. Peak impact force was typically reached at around 70 ms, so each trial recording consisted of 200 ms of data to ensure the peak was captured. It was assumed that the maximum vertical force, as measured by the load cells, and the maximum vertical linear acceleration, as measured by the tibial IMUs, occurred synchronously. This assumption was made based on previous studies that showed a strong correlation between GRF and linear acceleration [24]. Because both training and testing the validity of our models were needed, we separated our specimen pool into a training set (used in the development of our model; $N = 13$) and a testing set (used in the testing of model validity; $N = 3$).

2.2 Data Processing. Processing began with the removal of trials in which muscle forces and/or drop height or weight needed adjustment, where some component of the testing rig broke or malfunctioned, or where the CLS showed missing marker data (either "real" or "imaginary"). IMU data were then interpolated using the spline algorithm to match the frame rate of the sensors to the CLS. A fourth-order, zero-lag, low pass Butterworth filter was applied to the data from both the tibial and femoral IMUs (linear acceleration and angular velocity). Only the training set was utilized in determining the optimal frequencies. Optimum cutoff frequency for the Butterworth low-pass digital filter was obtained by applying Winter's method as used in previous work [36,37]. Average optimal cutoff frequencies were determined as: 9 Hz for the accelerometer ($X = 9$ Hz; $Y = 9$ Hz; $Z = 9$ Hz); 8 Hz for the gyroscope ($X = 8$ Hz; $Y = 8$ Hz; $Z = 8$ Hz) and were the same for the tibial and femoral IMU.

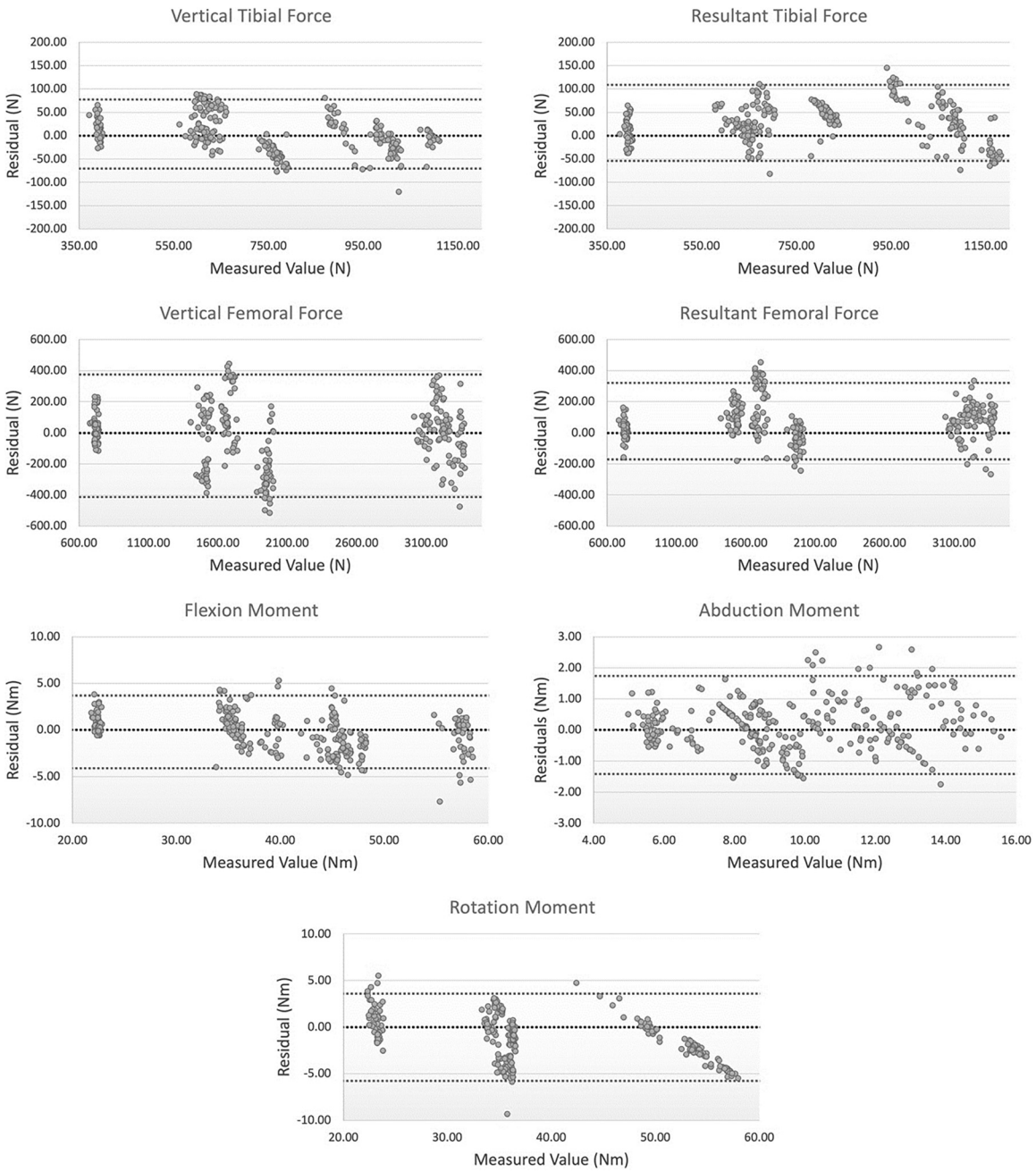

Fig. 3 Bland–Altman plots for each of the parameters for which the residual difference between predicted and measured is located on the *y*-axis and the measured value on the *x*-axis

Table 3 Results of model application for both training and testing sets across all seven metrics

Model	RMSE	R-squared	MSE	MAE
	Training results (5-fold cross-validation)			
Vertical tibial force	26.2 N	0.99	687	14.8 N
Vertical femoral force	102 N	0.97	10,505	70.9 N
Resultant tibial force	28.2 N	0.98	796	13.9 N
Resultant femoral force	80.4 N	0.98	6468	56.8 N
Flexion moment	1.51 N·m	0.99	2.28	0.85 N·m
Abduction moment	1.12 N·m	0.98	1.25	0.63 N·m
Rotational moment	1.48 N·m	0.97	2.12	0.84 N·m

Model	RMSE	R-squared	MAE	MAPE	NRMSE	LoA
			Testing results			
Vertical tibial force	37.9 N	0.97	29.4 N	4.42%	11.6%	(–70.6, 77.7) N
Vertical femoral force	202 N	0.95	159 N	9.14%	12.4%	(–413, 375) N
Resultant tibial force	49.6 N	0.97	40.4 N	5.34%	13.2%	(–53.9, 109) N
Resultant femoral force	145 N	0.98	111 N	6.18%	8.93%	(–171, 321) N
Flexion moment	1.99 N·m	0.97	1.55 N·m	3.97%	17.2%	(–4.08, 3.69) N·m
Abduction moment	0.82 N·m	0.93	0.64 N·m	6.77%	19.9%	(–1.41, 1.74) N·m
Rotation moment	2.62 N·m	0.95	2.07 N·m	5.42%	17.1%	(–5.77, 3.61) N·m

Root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), mean absolute percent error (MAPE), normalized root-mean-square error (NRMSE), and limits of agreement (LoA) are presented.

2.3 Model Development. We used the "regression learning" app developed by MATLAB to apply several types of machine learning algorithms to each metric (peak tibial or femoral forces; knee moments). To address the potential of overfitting and bias, a five-fold cross-validation was performed during each model's construction; this method is like the "leave-one-out" method but was applied five times on smaller subsets of the training set. Various adjustable modeling parameters (i.e., basis function, kernel function, isotropic nature of kernel function, number of layers, and activation function) for each model were tuned to obtain models with the lowest root-mean-square error (RMSE) values with respect to the training set. Following that, a Bayesian optimization with an "expected improvement per second" acquisition function was applied for 30 iterations to tune numerical parameters (i.e., kernel scale, sigma value, epsilon value, box constraint, and layer size). We determined the best model based on two criteria: normalized root-mean-square error (NRMSE) and Bland–Altman (BA) limits of agreement (LoAs) for the testing set. NRMSE was constructed as the RMSE value between predicted and measured values over the interquartile range of the measured value; these served as a measure of a model's accuracy as used elsewhere [27]. Based on the work of others NRMSE was limited to 20%, so only models that exhibited values at or below 20% were deemed accurate enough for further consideration [27]. LoAs were the measure of a model's reliability, and while often established a priori based on clinical/researcher specifications, in this study we utilized the 95% confidence interval (CI) of the residuals. Hence the smaller the 95% CI, the more reliable the model. The model with an acceptable level of accuracy and smallest LoAs was then determined as the best-fitting model. To further our analysis, BA plots were developed for descriptive analysis purposes to illustrate trends or bias that may be present. We recognize that our dependency on machine learning algorithms may preclude the development of a prediction equation (i.e., a simple model with defined variables and coefficients) and result in the "black-box" phenomenon. So, F-test graphs were used to determine the strength of the correlation between the parameter and predicted metric.

A confusion matrix was constructed to determine the specificity, precision, recall, and accuracy in correctly classifying "potentially injurious" from and "noninjurious" loading cycles. Two matrices were developed to compare the measured values (true class) to the predicted values (estimated class). The first compared events were classified based on the vertical femoral force (low risk: <3x BW; high risk: >3x BW). The second compared events were classified based on the rotation moment (low risk: <45.5 N·m; high risk: >45.5 N·m). These values were determined based on previous studies examining noncontact ACL injuries [7,38].

3 Results

Application of the "regression learning" app to each of the metrics generated a unique machine learning algorithm. These differed by: type (Gaussian process regression (GPR), neural network, or support vector machine (SVM)); basis function for GPR only (zero, constant, or linear); kernel function for GPR or SVM (rational quadratic, exponential; isotropic or nonisotropic); or number of layers for neural network only (single layer, bilayer) (Table 2). A GPR model had the lowest RSME for both the vertical and resultant tibial force and flexion and abduction moment, though each had different basis and kernel functions (Table 2). For the vertical tibial force, a GPR with constant basis function and rational quadratic kernel function was best, whereas a GPR with zero basis function and exponential kernel function was best for resultant tibial force. A GPR with a quadratic kernel and a constant basis function was also best for abduction moment predictions, though here the kernel was nonisotropic (i.e., not Euclidean distance dependent). The optimal model to predict knee flexion moment was a GPR with a nonisotropic 3/2 Matérn kernel function and zero-basis function.

Feed-forward neural networks were the best model type to predict both the vertical and resultant femoral forces, though with differing layer structures (Table 2). Vertical femoral force was best modeled through a single hidden layer, narrow neural network. The resultant femoral force was best modeled through a bilayer neural network that uses two hidden layers instead of one. In both cases, layer size was set to 10 nodes. Both neural networks' activation function was a rectified linear unit (ReLU). The rotation moment was best predicted using an SVM algorithm, more specifically a nonlinear SVM regression with a Gaussian kernel function (Table 2).

All force models developed exhibited a very strong level of fit (R^2 range: 0.95 to 0.98) and had NRMSEs below our pre-established conditions, with the resultant femoral force model having the lowest of these values (Table 2) (Fig. 2). There were no discernable trends as force increased in magnitude, illustrating that at any given measured value, a predicted value has consistent variability (Fig. 3). The resultant force model tended to overestimate force; this was evidenced by the upper bound condition being larger than the lower bound (Fig. 3). Vertical force models showed near equal tendencies for overestimations as underestimations. LoAs for the tibial models

fell within ±15% while the femoral force models exceeded 20% (Table 3).

Knee moment models all exhibited strong levels of fit (R^2 range: 0.93 to 0.97), and had NRMSE values below our pre-established condition, however, compared to the force models, the NRMSE values of the knee moment models were larger (Fig. 2 and Table 2). The BA plots showed varying degrees of bias and some trends. The knee flexion moment model tended to slightly underestimate the moment, and a linear slope was observed between the residuals and measured flexion moment; lower measured values tended to be overestimations, while higher values tended to be underestimations (Fig. 3). The knee abduction moment model had an overall tendency to overestimate the value, however, there was no discernable trend throughout the range of measured values; it should be noted however, that as the magnitude of the knee moment increased, variation also increased (Fig. 3). The knee transverse plane rotation moment model tended to slightly underestimate predicted values, and a negative linear relationship between the residual and measured value was observed (Fig. 3). The LoAs for all the moment models fell below ±20% (Table 3).

Classification of events as either potentially injurious or non-injurious was performed with a 99.7% accuracy for force classification and a 99.3% accuracy for rotation moment classification (Fig. 4). In both cases, recall was 100% with no instances of false negatives. Precision for force classification was 99.0% while that of the rotation moment classification was 97.9%. Specificity for the force classification was 99.5% and 99.0% when using the rotation moment classification (Fig. 4).

4 Discussion: Parameter Significance and Model Performance

In the introduction, we reviewed studies that utilized regression modeling techniques to establish correlations between data obtained from an IMU and metrics measured by a camera-based motion capture system. In this study, however, our goal was to develop models to predict those metrics. To that end, we used machine learning and considered all parameters measured by an IMU as well as mass to predict forces and knee moments during one-legged landing. We recognize that the added complexity of machine learning carries with it the benefits of increased accuracy at the cost of computational speed and potential training set dependency but given that our models are to be applied postcollection, computational speed is not a factor of concern. We applied our model to a testing set to examine their performance, thereby addressing the possibility of training set dependency. Finally, we examined whether we could accurately classify loading cycles as potentially injurious using information from a wearable sensor.

The vertical (x-axis) tibial linear acceleration was the largest contributing feature for both the tibial and femoral force models (Fig. 5). This was expected because of the relationship between force and acceleration. In the vertical and resultant tibia force models, other features that contributed included the flexion (z-axis) angular velocity of the tibia, x-axis femoral linear acceleration, and z-axis angular velocity of the femur. Z-axis angular velocity is correlated to ground reaction force during landing—an increase in angular velocity has been reported to decrease ground reaction force [24,39]. As mentioned in the Results, the vertical models showed near equal tendencies for overestimations as underestimations as opposed to the resultant models which tended to overestimate the predicted value. This may be due to the more complex nature of resultant forces; while vertical force is uniaxial, resultant forces are triaxial.

Several human subjects' studies have predicted GRF from the data obtained from an IMU using different techniques and sensors [27,40,41]. We used cadaveric specimens to rigidly fix the sensors to the segments thereby minimizing any soft tissue artifacts that are usually present when wearable sensors are fixed to human subjects. Even with no soft tissue, the predicted values had some error, indicating that this error may be caused by factors other than soft tissue movement. Thiel et al. correlated IMU data (triaxial linear acceleration) and stride count of the subject with force using multivariable linear regression models [40]. In that study, 15 subjects ran five 50 m sprints from a block start on an instrumented running track [40]. They constructed models for each individual subject, not a group, which may potentially limit the model's effectiveness for universal use on-field [40]. They reported moderate correlations between predicted and measured GRF with mean absolute percent error (MAPE) values ranging from 3.29% to 33.3% for each individual [40]. Although we measured tibia and femur force, not ground reaction force, in our study all force-related models had average MAPEs that were lower than these values. This may be the result of using machine learning or the mitigation of soft tissue artifact.

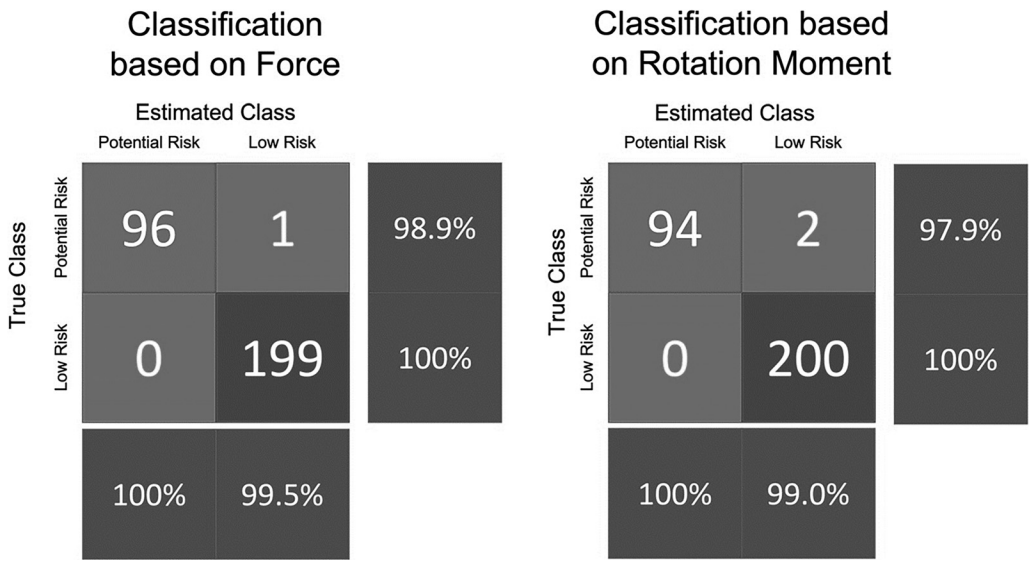

Fig. 4 Confusion matrices comparing the measured classification to the estimated one. The femoral measured force (true class) is on the left and compared to the estimated femoral force (estimated class). On the right is the measured rotation moment (true class) compared to estimated rotation moment (estimated class). Column and row normalization of instances of correctly classified events are presented either to the right (column) or bottom (row) of each matrix.

A study conducted by Neugebauer et al. reported strong correlations (similar to ours) in GRF ($R^2 = 0.97$) between predicted and measured values when applying mixed effect models to IMU data during gait [41]. They reported a MAPE of $5.2 \pm 1.6\%$ [41]. This study produced more accurate results than previously reported, possibly due to the added complexity of mixed effect models over generalized regression models. However, Neugebauer et al. only used linear accelerometer data; this does not consider potential angular movement contributions to forces. Angular rates were among the largest contributing factors to our models.

Lee et al. used machine learning to develop algorithms that predicted GRF as well as joint torques of the lower limbs, during normal walking using only feedforward neural networks. It was not clear whether other algorithms were considered [27]. In our study, a wide variety of learning algorithms were applied as we assumed different metrics may exhibit different relationships with the IMU data. Lee et al. observed a very strong fit between predicted and measured GRF ($R^2 = 0.93$), which is comparable to what we found [27]. They also presented NRMSE values, normalized to the amplitude of the output for each stance phase. In the case of average GRF, a NRMSE of 6.70% was reported which is lower than any NRMSE of our study [27]. Although, the lower value could be attributable to differing normalization techniques or the lack of an exclusive testing set, they used the leave-one-out method [27]. Because we had a larger set of data, we used a k-fold cross-validation [42].

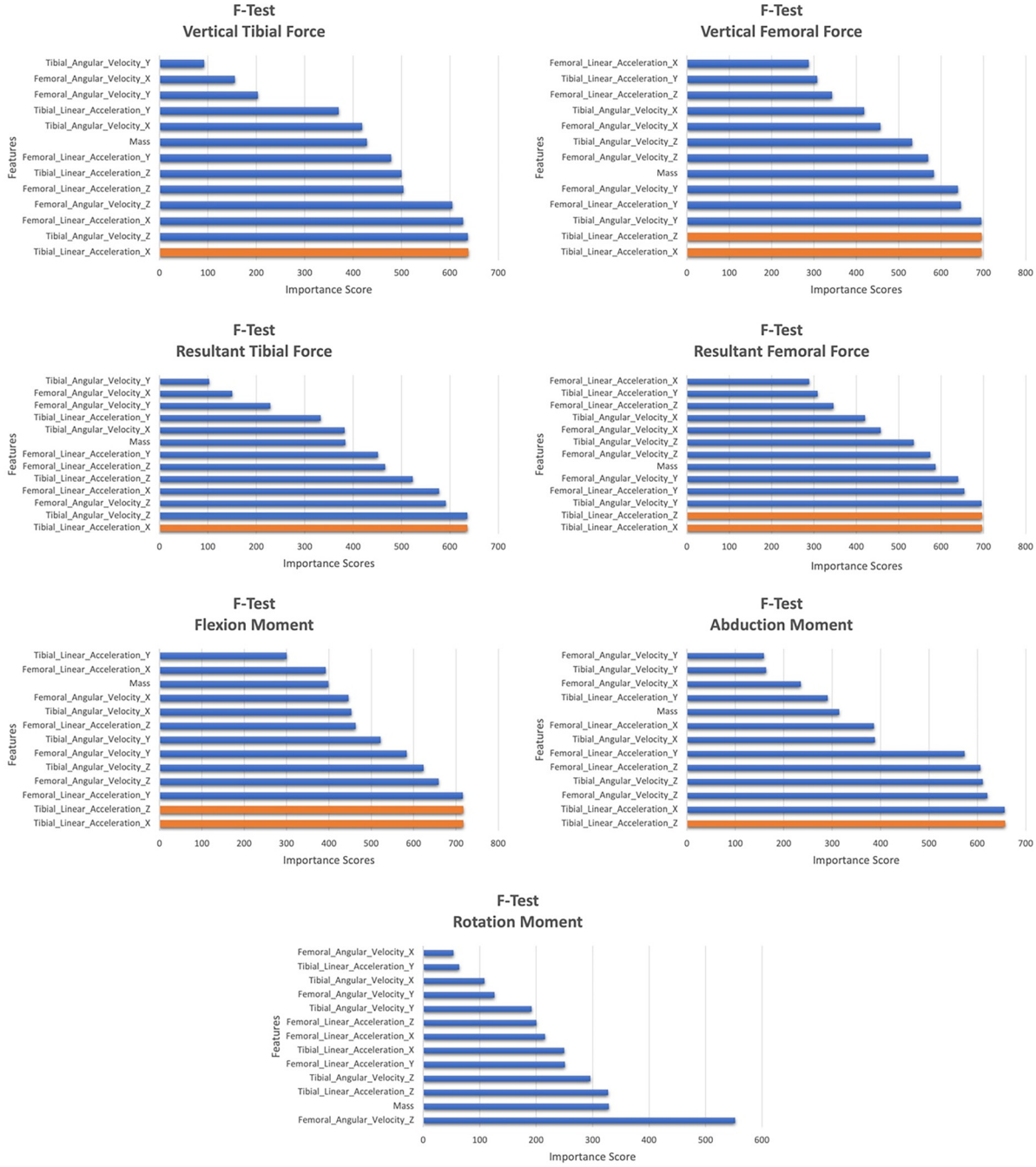

Fig. 5 F-test graphs used in determining how well correlated a variable was to a prediction model. On the *y*-axis are the differing features used in the construction of the model while on the *x*-axis are the importance scores. Larger scores indicate more significant features. Features in orange indicate the associated *p*-value was sufficiently small as to be considered zero.

The contributions of each IMU feature to prediction differed across the moment models. Flexion moment showed the least clear divide between those of more and those of lesser importance, however, the top three features were the *x*-axis and *z*-axis tibial linear accelerations and as well as the *y*-axis femoral linear acceleration (Fig. 5). This shows a potentially more dependent nature of flexion moments on tibial components. Abduction moment showed a clear divide between the upper six features and the lower seven, with *z*-axis and *x*-axis tibial linear accelerations, *x*-axis angular velocity of both the tibia and the femur, and the *z*-axis and y-axis femoral linear accelerations being the largest contributing features (Fig. 5). Unlike most other models, the abduction moment model was equally dependent on the data from the femur IMU as the tibia IMU. The greatest divide between the importance of the features was for the rotation moment model, where *y*-axis angular velocity of the femur was the largest contributing feature for prediction, having an *F*-score nearly twice that of the next most contributing feature, mass (Fig. 5). This observation was counterintuitive to what we initially theorized which was that the *x*-axis angular velocities of both segments would be the greatest contributing feature, however, both were ranked among the least contributing. This may be due to anatomical constraints of the knee that do not allow a significant amount of rotation.

As noted in the results, moment models exhibited larger NRMSE than those of the force models; this may be due to a larger range in force measurements, which in turn would mean an increased interquartile range value and ultimately a lower NRMSE value. Thus, while moment models all appear to have large NRMSE, it may be simply due to the nature of moments having smaller observed ranges. Of interest when discussing moments are the mean absolute error (MAEs). Moment differences between predicted and measured values less than 1 Nm may effectively be negligible for clinical relevance. Studies conducted by Shin et al., Levine et al., and Kiapour et al. have tested the strain response of the anterior cruciate ligament to both abduction and rotational moments, and have reported that knee abduction moments up to 51 N·m and internal rotation moments up to 30 N·m during noninjurious landing [16,43,44]. High-risk injurious moments can be up to three times those [45,46]. Therefore, while abduction moments illustrated the largest NRMSE, because the associated MAE was below 1 N·m, this may be clinically insignificant, as we would be able to differentiate between noninjurious and potentially "at-risk" moments.

Numerous studies have used IMU data in the development of prediction models for abduction and rotation moments, with varying success, though in most cases, low correlations were reported [27,33,47]. Konrath et al. created regression models using IMU data obtained from commonly performed rehabilitation activities for abduction moments. Three participants were fitted with two placed on the thigh and shank and asked to perform a stair ascent/descent as well as a sit-to-stand action. Strong, to very strong correlations between their predicted and measured values were reported which were comparable to ours [47]. They also reported very low NRSME values (NRMSE: SA: 0.01; SD: 0.014; StS: 0.006), though the direct comparison between our NRMSE and theirs cannot be done due to our differing normalization techniques [47]. Sit-to-stand is a slower action that is not as dynamic as a jump landing and may have also contributed to the low NRMSE values. Karatsidis et al. predicted rotational moments during walking of the knee using a full body internal motion capture (IMC) method [33]. Strong correlations between their predicted and measured values were reported ($R = 0.82$), which were slightly lower than that observed by our study [33]. The IMC system is made up of many individual IMUs and provides information on the kinematics of the entire body, rather than the limited view of a two IMU system. However, use of an IMC system on-field may not be practical in a real-world application.

Previous studies did not present a measure of reliability. Measuring and reporting the accuracy of a model is important, but arguably of equal importance is the model's reliability of measurement. Poor reliability would indicate predications are random in nature; while some instances may be highly accurate, others could be gross over/under estimations. Using an IMU to estimate whether landing techniques could possibly be injurious requires consistency. We also note that no study applied their model to data outside of a training set. Error presented through training set predictions will be smaller than the true error, and thus may be misleading with respect to the generalizability of the model to new points. Exclusive testing data to validate a model's accuracy/reliability is needed.

In recent years, IMUs have been used as tools in risk assessment and event classification, though this has been mostly limited to lifting tasks [48,49]. A study by Brandt et al. used a tri-axial accelerometer and a linear discriminate analysis algorithm to classify events as risky based on the Danish Working Environment Authority guidelines and reported event classification accuracy reaching 65.5% [49]. Donsis et al. also examined lifting tasks using IMUs and the NIOSH (National Institute of Occupational Safety and Health) standard to distinguish between "at risk" and "no risk" events [48]. They reported high accuracy (82.8%), sensitivity (84.8%), and specificity (80.9%) [48]. We demonstrated a 99.7% accuracy in classifying events as either potentially injurious or noninjurious, which is comparable.

4.1 Limitations. There are several limitations in this study including the use of cadaveric specimens. If the models developed during this stage were directly applied to human subjects, there would likely be errors between predicted and measured values due to noise pollution produced by motion of the soft tissue. We also acknowledge the potential over-tuning of our models to a specific action. Due to the limitations of the rig, we were only able to simulate a one-legged jump landing. Other actions, such as cutting and sudden decelerations have also been linked to noncontact ACL injuries [4]. More types of actions would be required to develop models robust enough to predict kinematics under any on-field condition. In addition, we acknowledge that risk assessment and classification of an event as either potentially injurious or not is not based on a single variable, but instead a combination of multiple variables. While in this work we focused on correctly assessing the risk based on force and rotational moments, we did so by treating each variable separately. In reality, risk assessment could be based on the combination of the force, rotational moment, abduction moment, and the orientation of knee at the time of loading. Future testing should focus on constructing a classification system that incorporates all these variables when assessing risk.

5 Conclusion

For these in vitro studies, each model demonstrated an acceptable level of accuracy with NRMSE values of less than 20% in a testing set independent of the data used for training the model. While several models exhibited some bias, most models exhibited LoAs below 20%. The femoral force models exhibited LoAs that exceeded 20%. Classification based on either force or rotation moment yielded high levels of accuracy, precision, specificity, and recall force correctly classifying whether an event was potentially injurious if repeated a sufficient number of times to cause ligament fatigue failure. Future studies should examine whether such models can be applied to in vivo movement data that will inevitably include unavoidable soft tissue movement artifacts.

Acknowledgment

The authors wish to thank Dr. Melanie Beaulieu for her assistance with data and Dr. Samir Rawashdeh for his assistance with the machine learning algorithms.

Funding Data

- Division of Chemical, Bioengineering, Environmental, and Transport Systems, National Science Foundation (Grant No. 2003434; Funder ID: 10.13039/100000146).

- National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health (Award No. AR054821; Funder ID: 10.13039/100000069).

Conflicts of Interest

The authors declare no conflict of interest.

Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request.

References

[1] Alghamdi, W., Alzahrani, A., Alsuwaydi, A., Alzahrani, A., Albaqqar, O., Fatani, M., and Alaidarous, H., 2017, "Prevalence of Cruciate Logaments Injury Among Physical Education Students of Umm Al-Qura University and the Relation Between the Dominant Body Side and Ligament Injury Side in Non-Contact Injury Type," Am. J. Med. Med. Sci., **7**(1), pp. 14–19.
[2] Saeed, I. O., 2018, "MRI Evaluation for Post-Traumatic Knee Joint Injuries," J. Nursing Health Sci., **7**(2), pp. 48–51.
[3] Hewett, T. E., Ford, K. R., and Myer, G. D., 2006, "Anterior Cruciate Ligament Injuries in Female Athletes: Part 2, A Meta-Analysis of Neuromuscular Interventions Aimed at Injury Prevention," Am. J. Sports Med., **34**(3), pp. 490–498.
[4] Hewett, T. E., Myer, G. D., and Ford, K. R., 2006, "Anterior Cruciate Ligament Injuries in Female Athletes: Part 1, Mechanisms and Risk Factors," Am. J. Sports Med., **34**(2), pp. 299–311.
[5] Myer, G. D., Paterno, M. V., Ford, K. R., Quatman, C. E., and Hewett, T. E., 2006, "Rehabilitation After Anterior Cruciate Ligament Reconstruction: Criteria-Based Progression Through the Return-to-Sport Phase," J. Orthop. Sports Phys. Ther., **36**(6), pp. 385–402.
[6] Paterno, M. V., Schmitt, L. C., Ford, K. R., Rauh, M. J., Myer, G. D., Huang, B., and Hewett, T. E., 2010, "Biomechanical Measures During Landing and Postural Stability Predict Second Anterior Cruciate Ligament Injury After Anterior Cruciate Ligament Reconstruction and Return to Sport," Am. J. Sports Med., **38**(10), pp. 1968–1978.
[7] Wojtys, E. M., Beaulieu, M. L., and Ashton-Miller, J. A., 2016, "New Perspectives on ACL Injury: On the Role of Repetitive Sub-Maximal Knee Loading in Causing ACL Fatigue Failure," J. Orthop. Res., **34**(12), pp. 2059–2068.
[8] Zitnay, J. L., Jung, G. S., Lin, A. H., Qin, Z., Li, Y., Yu, S. M., Buehler, M. J., and Weiss, J. A., 2020, "Accumulation of Collagen Molecular Unfolding is the Mechanism of Cyclic Fatigue Damage and Failure in Collagenous Tissues," Sci. Adv., **6**(35), p. eaba2795.
[9] Kim, J., Baek, S., Schlecht, S. H., Beaulieu, M. L., Bussau, L., Chen, J., et al., 2022, "Anterior Cruciate Ligament Microfatigue Damage Detected by Collagen Autofluorescence In Situ," J. Exp. Orthop., **9**(1), p. 74.
[10] Putera, K. H., Kim, J., Baek, S. Y., Schlecht, S. H., Beaulieu, M. L., Haritos, V., Arruda, E. M., Ashton-Miller, J. A., Wojtys, E. M., and Banaszak Holl, M. M., 2023, "Fatigue-Driven Compliance Increase and Collagen Unravelling in Mechanically Tested Anterior Cruciate Ligament," Commun. Biol., **6**(1), p. 564.
[11] Griffin, L. Y., Agel, J., Albohm, M. J., Arendt, E. A., Dick, R. W., Garrett, W. E., Garrick, J. G., et al., 2000, "Noncontact Anterior Cruciate Ligament Injuries: Risk Factors and Prevention Strategies," J. Am. Acad. Orthop. Surg., **8**(3), pp. 141–150.
[12] Olsen, O.-E., Myklebust, G., Engebretsen, L., and Bahr, R., 2004, "Injury Mechanisms for Anterior Cruciate Ligament Injuries in Team Handball: A Systematic Video Analysis," Am. J. Sports Med., **32**(4), pp. 1002–1012.
[13] Hollis, J. M., Takai, S., Adams, D. J., Horibe, S., and Woo, S. L.-Y., 1991, "The Effects of Knee Motion and External Loading on the Length of the Anterior Cruciate Ligament (ACL): A Kinematic Study," ASME J. Biomech. Eng., **113**(2), pp. 208–214.
[14] Markolf, K. L., Burchfield, D. M., Shapiro, M. M., Shepard, M. F., Finerman, G. A. M., and Slauterbeck, J. L., 1995, "Combined Knee Loading States That Generate High Anterior Cruciate Ligament Forces," J. Orthop. Res., **13**(6), pp. 930–935.
[15] Yu, B., and Garrett, W. E., 2007, "Mechanisms of Non-Contact ACL Injuries," Br. J. Sports Med., **41**(Suppl 1), pp. i47–i51.
[16] Shin, C. S., Chaudhari, A. M., and Andriacchi, T. P., 2011, "Valgus Plus Internal Rotation Moments Increase Anterior Cruciate Ligament Strain More Than Either Alone," Med. Sci. Sports Exer., **43**(8), pp. 1484–1491.
[17] Anderson, C. J., Ziegler, C. G., Wijdicks, C. A., Engebretsen, L., and LaPrade, R. F., 2012, "Arthroscopically Pertinent Anatomy of the Anterolateral and Posteromedial Bundles of the Posterior Cruciate Ligament," J. Bone Jt. Surg. Am., **94**(21), pp. 1936–1945.
[18] Arif, M., and Kattan, A., 2015, "Physical Activities Monitoring Using Wearable Acceleration Sensors Attached to the Body," PLoS One, **10**(7), p. e0130851.
[19] Cutti, A. G., Ferrari, A., Garofalo, P., Raggi, M., Cappello, A., and Ferrari, A., 2010, "Outwalk': A Protocol for Clinical Gait Analysis Based on Inertial and Magnetic Sensors," Med. Biol. Eng. Comput., **48**(1), pp. 17–25.
[20] Favre, J., Crevoisier, X., Jolles, B. M., and Aminian, K., 2010, "Evaluation of a Mixed Approach Combining Stationary and Wearable Systems to Monitor Gait Over Long Distance," J. Biomech., **43**(11), pp. 2196–2202.
[21] Ferrari, A., Cutti, A. G., Garofalo, P., Raggi, M., Heijboer, M., Cappello, A., and Davalli, A., 2010, "First In Vivo Assessment of "Outwalk": A Novel Protocol for Clinical Gait Analysis Based on Inertial and Magnetic Sensors," Med. Biol. Eng. Comput., **48**(1), pp. 1–15.
[22] Lin, J. F., and Kulić, D., 2012, "Human Pose Recovery Using Wireless Inertial Measurement Units," Physiol. Meas., **33**(12), pp. 2099–2115.
[23] Picerno, P., Cereatti, A., and Cappozzo, A., 2008, "Joint Kinematics Estimate Using Wearable Inertial and Magnetic Sensing Modules," Gait Posture, **28**(4), pp. 588–595.
[24] Elvin, N. G., Elvin, A. A., and Arnoczky, S. P., 2007, "Correlation Between Ground Reaction Force and Tibial Acceleration in Vertical Jumping," J. Appl. Biomech., **23**(3), pp. 180–189.
[25] Gurchiek, R. D., McGinnis, R. S., Needle, A. R., McBride, J. M., and van Werkhoven, H., 2017, "The Use of a Single Inertial Sensor to Estimate 3-Dimensional Ground Reaction Force During Accelerative Running Tasks," J. Biomech., **61**, pp. 263–268.
[26] Meyer, U., Ernst, D., Schott, S., Riera, C., Hattendorf, J., Romkes, J., Granacher, U., Göpfert, B., and Kriemler, S., 2015, "Validation of Two Accelerometers to Determine Mechanical Loading of Physical Activities in Children," J. Sports Sci., **33**(16), pp. 1702–1709.
[27] Lee, M., and Park, S., 2020, "Estimation of Three-Dimensional Lower Limb Kinetics Data During Walking Using Machine Learning From a Single IMU Attached to the Sacrum," Sensors (Basel), **20**(21), p. 6277.
[28] Alcantara, R. S., Edwards, W. B., Millet, G. Y., Grabowski, A. M., 2022, "Predicting Continuous Ground Reaction Forces From Accelerometers During Uphill and Downhill Running: A Recurrent Neural Network Solution," PeerJ, **10**, p. e12752.
[29] Tedesco, S., Alfieri, D., Perez-Valero, E., Komaris, D-S., Jordan, L., Belcastro, M., Barton, J., Hennessy, L., and O'Flynn, B., 2021, "A Wearable System for the Estimation of Performance-Related Metrics During Running and Jumping Tasks.," Appl. Sci., **11**(11), p. 5258.
[30] Liang, W., Wang, F., Fan, A., Zhao, W., Yao, W., and Yang, P., 2023, "Extended Application of Inertial Measurement Units in Biomechanics: From Activity Recognition to Force Estimation," Sensors (Basel), **23**(9), p. 4229.
[31] McLean, S. G., Lipfert, S. W., and van den Bogert, A. J., 2004, "Effect of Gender and Defensive Opponent on the Biomechanics of Sidestep Cutting," Med. Sci. Sports Exer., **36**(6), pp. 1008–1016.
[32] Sigward, S. M., and Powers, C. M., 2007, "Loading Characteristics of Females Exhibiting Excessive Valgus Moments During Cutting," Clin. Biomech. (Bristol, Avon), **22**(7), pp. 827–833.
[33] Karatsidis, A., Jung, M. K., Schepers, M., Bellusci, G., De Zee, M., Veltink, P. H., and Andersen, M. S., 2018, "Predicting Kinetics Using Musculoskeletal Modeling and Inertial Motion Capture," arxiv:1801.01668.
[34] Lipps, D. B., Wojtys, E. M., and Ashton-Miller, J. A., 2013, "Anterior Cruciate Ligament Fatigue Failures in Knees Subjected to Repeated Simulated Pivot Landings," Am. J. Sports Med., **41**(5), pp. 1058–1066.
[35] Lipps, D. B., Oh, Y. K., Ashton-Miller, J. A., and Wojtys, E. M., 2014, "Effect of Increased Quadriceps Tensile Stiffness on Peak Anterior Cruciate Ligament Strain During a Simulated Pivot Landing," J. Orthop. Res., **32**(3), pp. 423–430.
[36] Yu, B., Gabriel, D., Noble, L., and An, K.-N., 1999, "Estimate of the Optimum Cutoff Frequency for the Butterworth Low-Pass Digital Filter," J. Appl. Biomech., **15**(3), pp. 318–329.
[37] Ajdaroski, M., Ashton-Miller, J. A., Baek, S. Y., and Shahshahani, P. M., et al., 2021, "Testing a Quaternion Conversion Method to Determine Human 3D Tibiofemoral Angles During an In Vitro Simulated Jump Landing," ASME J. Biomech. Eng., **144**(4), p. 041002.
[38] Bates, N. A., Schilaty, N. D., Nagelli, C. V., Krych, A. J., and Hewett, T. E., 2019, "Multiplanar Loading of the Knee and Its Influence on Anterior Cruciate Ligament and Medial Collateral Ligament Strain During Simulated Landings and Noncontact Tears," Am. J. Sports Med., **47**(8), pp. 1844–1853.
[39] Yu, B., Lin, C. F., and Garrett, W. E., 2006, "Lower Extremity Biomechanics During the Landing of a Stop-Jump Task," Clin. Biomech. (Bristol, Avon), **21**(3), pp. 297–305.
[40] Thiel, D. V., Shepherd, J., Espinosa, H. G., Kenny, M., Fischer, K., Worsey, M., Matsuo, A., and Wada, T., 2018, "Predicting Ground Reaction Forces in Sprint Running Using a Shank Mounted Inertial Measurement Unit," Proceedings, **2**(6), p. 199.
[41] Neugebauer, J. M., Hawkins, D. A., and Beckett, L., 2012, "Estimating Youth Locomotion Ground Reaction Forces Using an Accelerometer-Based Activity Monitor," PLoS One, **7**(10), p. e48182.
[42] Wong, T.-T., 2015, "Performance Evaluation of Classification Algorithms by K-Fold and Leave-One-Out Cross Validation," Pattern Recogn., **48**(9), pp. 2839–2846.
[43] Levine, J. W., Kiapour, A. M., Quatman, C. E., Wordeman, S. C., Goel, V. K., Hewett, T. E., and Demetropoulos, C. K., 2013, "Clinically Relevant Injury Patterns After an Anterior Cruciate Ligament Injury Provide Insight Into Injury Mechanisms," Am. J. Sports Med., **41**(2), pp. 385–395.
[44] Kiapour, A. M., Demetropoulos, C. K., Kiapour, A., Quatman, C. E., Wordeman, S. C., Goel, V. K., and Hewett, T. E., 2016, "Strain Response of the Anterior Cruciate Ligament to Uniplanar and Multiplanar Loads During Simulated Landings: Implications for Injury Mechanism," Am. J. Sports Med., **44**(8), pp. 2087–2096.
[45] Besier, T. F., Lloyd, D. G., Cochrane, J. L., and Ackland, T. R., 2001, "External Loading of the Knee Joint During Running and Cutting Maneuvers," Med. Sci. Sports Exer., **33**(7), pp. 1168–1175.

[46] McLean, S. G., Huang, X., Su, A., and van den Bogert, A. J., 2004, "Sagittal Plane Biomechanics Cannot Injure the ACL During Sidestep Cutting," Clin. Biomech. (Bristol, Avon), **19**(8), pp. 828–838.
[47] Konrath, J., Karatsidis, A., Schepers, H., Bellusci, G., de Zee, M., and Andersen, M., 2019, "Estimation of the Knee Adduction Moment and Joint Contact Force During Daily Living Activities Using Inertial Motion Capture," Sensors (Basel), **19**(7), p. 1681.
[48] Donisi, L., Cesarelli, G., Capodaglio, E., Panigazzi, M., D'Addio, G., Cesarelli, M., and Amato, F., 2022, "A Logistic Regression Model for Biomechanical Risk Classification in Lifting Tasks," Diagnostics (Basel), **12**(11), p. 2624.
[49] Brandt, M., Madeleine, P., Samani, A., Jakobsen, M. D., Skals, S., Vinstrup, J., and Andersen, L. L., 2018, "Accuracy of Identification of Low or High Risk Lifting During Standardised Lifting Situations," Ergonomics, **61**(5), pp. 710–719.