Abstract—Early identification of motion disparities in Anterior Cruciate Ligament reconstructed (ACL-R) athletes may better post-operative decision making when returning athletes to sport. Existing return to play assessments consist of assessments of muscle strength, functional tasks, patient-reported outcomes, and 3D coordinate tracking. However, these methods primarily depend on the medical provider’s intuition to release them to participate in an unrestricted activity after ACL-R that may cause reinjury or long-term impacts. This study proposes a wearable sensor-based system that helps track athlete rehabilitation progress and return to sport decision making. For this, we capture gait data from 89 ACL-R athletes during their walking and jogging trials. The raw gyroscope data collected from this system is used to extract causal features based on Nolte’s phase slope index. Features extracted from this study are used to develop computational models that classify ACL-R athletes based on their reconstructed knee during two visits (3-6 months & 9 months) post ACL-R surgery. The classifier’s performance degradation in detecting ACL-R athletes injured knee during multiple visits supports athletic trainers and physicians’ decision-making process to confirm an athlete’s safe return to sport.

Clinical Relevance—This study develops computational models based on causal analysis of gait data to support athletic trainers and medical practitioners’ decision to return athletes to sport post ACL-R surgery.

I. INTRODUCTION

An Anterior Cruciate Ligament Reconstruction (ACL-R) is a common surgery among high-level athletes, with approximately 250,000 surgeries in the U.S. per year [1]. Athletes elect to have ACL-R to return to prior activity levels; however, only 55% of competitive athletes return to these prior levels of sport [2]. Accompanying low rates of returning to the sport, these patients also experience poor subjective function, higher risks of subsequent injury, and greater risk of post-traumatic osteoarthritis [3]. These detrimental outcomes acutely following ACL-R and returning to activity may suggest that current return to sport assessment may not accurately identify functional deficits within the patient.

The current return to sport assessment consists of muscle strength, dynamic functional tasks, postural stability, and a validation questionnaire to track patient progress [4][5]. These studies show a lack of agreement in setting appropriate criteria that can be used to release athletes to unrestricted physical activity after ACL-R [6][7].

Clinical studies in kinesiology reported that the post-surgery variations in knee kinematics provide significant insights into the long-term consequences of an ACL-R [3][8]. A study by Miyazaki et al. reported an increase in knee varus moment by one percent contributed to a 6-factor growth in osteoarthritis progression [9]. Butler et al. [10] observed a 21 percent increase in knee varus moment while walking in subjects after ACL rupture compared to healthy individuals. These variations in knee moments increase as time progresses. A study also showed that the knee and hip level sagittal lower extremity kinetic adaptations are reported following ACL-R [11]. 6-12 months after ACL-R, individuals’ walking gait/pattern altered compared to a healthy group. These variations include reduced knee extension moments and reduced knee flexion [11]. This reduced knee extension and knee flexion angles are associated with weakness in the quadriceps up to a year post ACL-R [11].

The earlier research insights that specified gait variations postACL-R are used as the basis for our current study. This work hypothesizes that as time progresses and athletes get back to their regular fitness, the specific variations found in gait right after ACL-R surgery will slowly disappear. These variations found in an individual’s gait post ACL-R can be quantified to develop a methodology that supports tracking athletes’ rehabilitation progress. To achieve this, the current work develops a methodology that utilizes computational modeling methods that identify gait variations in signals captured by inertial sensors to classify the ACL-R knee of an athlete.

II. RELATED WORK

The recent emergence of inertial sensors in tracking human gait provides more precise and objective observations. Investigating gait data provides kinetic and kinematic information on individuals’ functional motor features that help define outcome evaluations and provide therapeutic interventions. Current gait analysis tools like force plates and stereophotogrammetry provide precise and quality data related to gait kinematics [14].

Standard tools also have several drawbacks, such as setup time, cost, and range, as they are confined to camera defined Virginia, Charlottesville, Virginia, 22908, USA (email: joehart@virginia.edu)

3 J. Lach is with the Department of Electrical & Computer Engineering, School of Engineering & Applied Science, George Washington University, Washington D.C., 20052, USA (email: jlach@gwu.edu)
space [14]. On the contrary, inertial sensors can be used to analyze gait data outside of laboratories and can be accessed by the researcher and clinicians regularly [11][14][15]. On the other hand, the inertial sensors are lower in cost and can be used in any space. The increase in the adoption of inertial sensors supports an extensive collection of gait data by clinicians and researchers to enhance rehabilitation methods and validate the research.

Earlier inertial sensors research in gait analysis focused on spatial and temporal features of gait data coming from the patient population to identify gait patterns, the pathologies in these studies are characterized by substantial gait deviations [14]. Observed differences showed changes in step length of 0.6 m for stroke patients [16] and 0.5 m for patients with Parkinson's disease [17]. Injuries like knee osteoarthritis and ACL-R alter angular kinematics and kinetics of joints instead of spatial and temporal deviations that can lead to joint impacting diseases like osteoarthritis.

Current research investigating inertial sensor data for angular kinematics is not highly accurate and clinically useful [14]. Error rates in these techniques vary significantly between a human knee and a robotic knee due to sensor placement and orientation issues [18]. The reported error rates in these studies make them suitable for identifying gait patterns with large deviations. Some abnormal gait patterns changes related to ACL-R have subtle variations that need to be identified from the gait data [14].

Our earlier study [11] used a causality-based approach that captures the causal interactions between different body parts that support better modeling. Nolte et al. [19] formalized the causal effects between different pairs of signals and named the interaction between signals as a phase slope index (PSI). Gong et al. [20] successfully applied this causal-based approach to quantify the interactions between body parts in multiple sclerosis-based patients. Inspired by this work [20], we extracted causal features from ACL-R motion data captured by wearable sensors to develop computational models that can classify athletes' ACL-R knee [11][12].

The promising results from this study [11] encourage us to extensively analyze causal features and develop a computational model-based rehabilitation method to track the subject's progress post ACL-R. This current study develops a rehabilitation tracking method by capturing gait data from ACL-R subjects during two visits (first visit 3-6 months post ACL-R and second visit 9 months post ACL-R). The gait data captured during the first visit is used to extract causal features and develop computational models that accurately classify the subject's ACL-R knee (Left (L) Vs. Right (R)). The models validated and trained on first visit data are applied on second visit data to track the prediction capability and confidence. The reduction in classification confidence of computational models during the second visit acts as a supporting indicator for subjects' return to play decision making.

III. METHODOLOGY

This section focuses on participant recruitment, data collection protocol, signal processing methods, and computational models to classify ACL-R individuals. Fig. 1 shows the proposed approach that starts with inertial measurement collection, extracts causal features, and then develops computational models for ACL-R classification.

A. Participant Demographics

A total of 89 participants with ACL-R knee were recruited for this study. Of these 89 participants, 8 participants also completed their visit 2 for data extraction. Patient demographics can be found in Table I. All ACL-R participants were referred from a single academic orthopedic clinic, and their first visit for data collection is scheduled approximately six months after surgery, and the second visit is 9 months post-surgery. Even though it makes sense to collect and add data before surgery, there is no realistic possibility as it is hard to predict which athlete will get injured or even from injured

![Fig. 2. Placement of inertial sensors on participant (Left) and data collection setting (Right)](image-url)

![Fig. 1. Proposed method to classify ACL-R subjects.](image-url)
athlete due to injury concerns. All participants obtained a primary, isolated ACL-R with no surgical complications. The university's institutional review board approved this study, and all participants provided written informed consent.

B. Gait Data from Inertial Sensors
To measure gait data, we adopt five inertial monitoring units developed by Shimmer Sensing. These IMU units are equipped with a tri-axial accelerometer, gyroscope, and magnetometer. Two sensors are placed on the distal shanks, two on the distal forearm, and one on the posterior sacrum, as shown in Fig. 2. Once the participants are equipped with these sensors, we inform them to walk for 5 minutes at a speed of 3 mph and jog for 3 minutes at a speed of 6 mph on a treadmill, as shown in Fig. 2. The data is collected at a 128 Hz sampling rate to attain maximum synchronization that supports complete body motion capture during walking and jogging trials.

C. Feature extraction from gait data
This study's primary focus is to extract causal features from gait data to develop computational models that classify reconstructed knee during two consecutive visits in one year.

1) Causal Features from Gait: Existing causality algorithms such as Granger causality and phase slope index (PSI) have strict restrictions on input signal stationary properties. Prior research in this domain identified that accelerometer data is merged with multiple artifacts that make it challenging to meet all the causality algorithms' stationary requirements [20]. On the contrary, the data captured by the gyroscope meets the stationary requirements of a causal algorithm. In line with our earlier studies [11][12], we extracted causal features every 6 seconds from the gyroscope data collected during walking and jogging trials. To do this, we first segment the data into 6-second subsets and calculate the PSI. As it is challenging to calculate causal relationships in real-world datasets that consist of confounding relations and time lags, PSI is a proven method to formalize causal relations between signals from different sensors. PSI's central concept is that the cause comes before the effect in time, which provides a correlation between the slope of signals and influences their directions [19][21]. The slope mainly reflects the cross-spectrum that occurs between signals [20]. The PSI values are arranged in the form of a causal matrix, as shown in Fig. 1. This study's causal features are extracted from the 3*3 matrix between the left ankle (LA) and right ankle (RA) as seen in Fig 1. This matrix is flattened out, and each value is named as features starting X1 to X9. These features act as the input for computation models to discriminate ACL-R knee.

D. Datasets
Temporal changes in gait data collected during walking and jogging are identified in individuals post ACL-R. To amplify the temporal changes from causality data, we use a moving average method that generates averaged individuals' causal features based on a sliding window method. A moving average method for increasing temporal resolution helps classification algorithms in learning from a block of local data points. As there is no set principle for choosing the window length and stride length, we try two different window lengths of 0.5 minute and 1 minute with a stride length of 0.1 minutes in this work. The nine causal features were transformed based on these moving averages to provide uniform samples for all individuals.

E. Computational Models
Our earlier study [11][13] showed promising performances in classifying ACL-R knee (Left Vs. Right) using neural network algorithms. This work only focuses on Left Vs. Right ACL prediction instead of Injured Vs. Uninjured due to data collection constraints. Predicting Left Vs. Right ACL-R is an acceptable approach as the focus of this study is to apply models longitudinally and analyze rehabilitation based on athlete recovery that can be interpreted from algorithm misclassification. Encouraged by the previous findings, we apply four deep learning methods that capitalize on spatial and temporal characteristics. This work adopts a dropout layer after each connected layer to avoid overfitting convolution neural networks (CNN) and recurrent neural networks (RNN/LSTM/GRU). The deep neural models are trained using mini-batches with a size of 32. All four models1 are trained using a multitude of epochs as high variations will make models either under or overfit the training data. This work also uses binary cross-entropy for classification and adam optimizer for network optimization.

We adopt a five-fold cross-validation process to validate the developed models that divide the data set into five sub folds. Four folds are used for training the model, and one fold is used for testing the model. This process repeats until the algorithm classifies all the observations into one of the two categories (Left Vs. Right Knee). The average of performances from five test folds was taken to evaluate the models. Once the model with high performance is observed, this work trains the best performing model on whole visit 1 data and then make predictions on visit 2 data to evaluate the classification confidence. The next section in this work reports the performance and confidence of algorithms.

IV. RESULTS
The results section in this study is divided into two subsections. The first subsection reports the four deep learning algorithms' classification performances on visit 1 and visit 2 multi-sample subject data. Each algorithm's discriminative capabilities are evaluated based on three metrics: Area under the curve (AUC), Accuracy, and Cohen's Kappa. The best algorithm is decided based on the tradeoff between AUC and kappa values. The second subsection shows the confidence of the best performing algorithm observed in the first subsection to predict visit 1 and visit 2 data.

A. Visit 1 & Visit 2 Classification Performance
In this analysis, we only report the best results from three different moving average windows tested. Based on the internal performance metrics evaluation, we observe that the data extracted from a moving average of 1 minute improved models' performance compared to raw and 0.5-minute

---

1 https://github.com/vmand4/Gait_DL_Models_EMBC
window data. Tables II and III below represent the performance of four deep learning models developed on 1 minute moving average gait causal data collected during visit 1 and visit 2, respectively. From the table I, we can observe that the convolution neural networks validated on visit 1 data performed better than the other three algorithms developed in this study. One reason for this is CNN's ability to characterize both spatial and temporal characteristics present in the input features. From Table III, we can observe that the classification performance of algorithms trained on visit 1 to predict data captured during visit 2 for 8 subjects reduced drastically. The CNN algorithm that classified ACL-R limb with 0.95 and 0.96 AUC during visit 1 dropped to 0.55 and 0.47 for walking and jogging data during visit 2. This reduction in performances prompted us to understand the confidence of algorithms in classifying each subject.

**TABLE II: MODEL PERFORMANCE FOR LEFT VS RIGHT ACL-R KNEE PREDICTION FOR VISIT 1 DATA**

<table>
<thead>
<tr>
<th>ML Algorithm</th>
<th>Walk</th>
<th>Jog</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC</td>
<td>Kappa</td>
</tr>
<tr>
<td>CNN</td>
<td>0.95</td>
<td>0.90</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.90</td>
<td>0.79</td>
</tr>
<tr>
<td>GRU</td>
<td>0.92</td>
<td>0.85</td>
</tr>
<tr>
<td>Simple RNN</td>
<td>0.94</td>
<td>0.89</td>
</tr>
</tbody>
</table>

**TABLE III: MODEL PERFORMANCE FOR LEFT VS RIGHT ACL-R KNEE PREDICTION FOR VISIT 2 DATA**

<table>
<thead>
<tr>
<th>ML Algorithm</th>
<th>Walk</th>
<th>Jog</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC</td>
<td>Kappa</td>
</tr>
<tr>
<td>CNN</td>
<td>0.55</td>
<td>0.091</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.52</td>
<td>0.043</td>
</tr>
<tr>
<td>GRU</td>
<td>0.50</td>
<td>-0.01</td>
</tr>
<tr>
<td>Simple RNN</td>
<td>0.53</td>
<td>0.062</td>
</tr>
</tbody>
</table>

**B. Visit 1 & Visit 2 Prediction Confidence**

The above subsection's algorithms’ performances are based on multi-sample subject data that consists of multiple prediction outcomes for each subject. As the confidence need to be calculated per subject rather than per sample, this work developed a confidence based approach. To achieve this, we capture the prediction made by algorithm on all samples in the data set and group the predictions based on each subject's identity value. Once the predictions are grouped, we calculate the count of all correct predictions and divide it by the total number of samples present for that subject. The outcome will be a confidence value between 0 and 1 (0 to 100 percent).

In order to understand the confidence of the algorithm in predicting each subject's ACL-R knee, this work developed a box plot for both visit 1 and visit 2 confidence values. Fig. 3 shows that the algorithm confidences in predicting the ACL-R knee of subjects during visit 2 reduced by a huge amount compared to visit 1. This finding is in line with our hypothesis that as the time progresses and the subject is getting back to their normal health condition, the specific variation found post ACL-R might disappear, and the data distributions will vary. These data distribution variations will make it harder for the ML algorithms to predict the original impacted knee.

**Fig. 3. Algorithm Confidences in classifying ACL-R (Left Vs Right) during Visit 1 and Visit 2**

**V. DISCUSSION**

Return to sport decision-making post ACL-R surgery without future complications and reinjury has been a huge challenge. This study focuses on developing a computational modeling based rehabilitation tracking of athletes using the causal features extracted from their gait data during two consecutive visits scheduled at 6 months and 9 months post ACL-R. The hypothesis is that computational models' ability to classify ACL-R knee reduces as time progresses, and the subject returns to their regular activity.

In general, computational models can predict well when the test data comes from a similar distribution of the data they were trained on. We use this fundamental concept in modeling to track rehabilitation progress. Earlier research indicated that gait data in athletes post ACL-R have identifiable variations, and these variations will disappear as the athlete recovers from ACL-R. Based on the modeling fundamental discussed above, the algorithm, when validated on first visit gait data, performs well in classifying ACL-R knee. It can capture the specific variations found post ACL-R. As time progresses and the subject recovers from ACL-R, gait data's variations start to disappear and reduce the model performance. This phenomenon was tested using a two sample Kolmogorov-Smirnov test that identifies if the data collected during visit 1 and visit 2 have similar distributions. We observed that the p-values are in the range of 0 to 0.092 for all 9 features from this nonparametric test as shown in Table IV. There are only two features with a p-value greater than 0.05, but they are still less than 0.092. As the p-values are very low, we can determine that the distribution of gait data collected during visit 1 and visit 2 varies. This phenomenon is also represented in the huge drop of performance metrics and confidence between visit 1 and visit 2, as shown in table II, III, and fig. 3. These interesting
findings from causal data also prompted us to study their interaction with domain variables.

### Table IV: P values of features after Kolmogorov-Smirnov test on visit 1 and visit 2 data

<table>
<thead>
<tr>
<th>Features</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>X2, X3, X4, X5, X6, X8, X9</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>X1 &amp; X7</td>
<td>&gt;0.05 &amp; &lt; 0.092</td>
</tr>
</tbody>
</table>

To understand the interactions between causal and domain-related features, we calculated the Pearson correlation coefficient between domain data variables collected using International Knee Documentation Committee (IKDC), Knee Injury and Osteoarthritis Outcome Score (KOOS), strength, symmetry, Tampa, and gait causal features. The correlation results show that the domain data is less correlated with gait causal features. This low correlation is specifically true in the case of variables from questionnaire data. One reason might be the nature of tasks being performed. Adopting causal-based systems is less intrusive and mitigates bias in the self-questionnaire to track rehabilitation. This work's implications help both trainers and athletes track the rehabilitation process and reduce the possibility of reinjury. Based on an algorithm's confidence value in detecting ACL-R individuals reconstructed knee, trainers and physicians can assess an athlete's recovery progress. This confidence value supports developing a personalized recovery treatment that helps athletes return to sport quickly and with a lower reinjury risk.

There are some potential limitations in the current study that will be addressed in our future research study. First, this study's data collection is done in a controlled environment where athletes walk and jog on a treadmill. It is crucial to study these algorithms' performance from the data collected during real practice sessions on training grounds. In addition to this, we are also working with domain experts to set a feasible confidence value that supports a return to sport decision-making. Moreover, a longitudinal study with more subjects coming for a visit 2 is in the works. This study's final opportunity is to develop a decision support system that considers domain-specific variables and sensor-based gait data to predict athlete return to sport accurately.

### References


