

CausalGait: Towards Personalized Gait Interpretation via Causality-Informed Knowledge Graphs

Mingrui Ma
mma36@illinois.edu
University of Illinois
Urbana-Champaign
Champaign, Illinois, USA

Muyu Yang
muyuy2@illinois.edu
University of Illinois
Urbana-Champaign
Champaign, Illinois, USA

Kylie Clewes
KClewes@stanfordchildrens.org
Stanford Medicine Children's
Health
Palo Alto, California, USA

Evan Lowe
elowe@stanfordchildrens.org
Stanford Medicine Children's
Health
Palo Alto, California, USA

Jessica Rose
jessica.rose@stanford.edu
Stanford University
Stanford, California, USA

Yiwen Dong
yiwen@illinois.edu
University of Illinois
Urbana-Champaign
Champaign, Illinois, USA

Abstract

Clinical gait interpretation is important to guide strategic intervention for neuromuscular disorders such as Cerebral Palsy (CP). Existing gait interpretation relies on either fixed clinical rules or population-based data-driven models to link gait observations with the underlying neuromuscular impairments, which can lead to one-size-fits-all intervention strategies with limited interpretability for a specific person. To close this gap, we introduce *CausalGait*, a framework that integrates clinical rules and measured gait data into personalized causal knowledge graphs, aiming to assist medical experts with more strategic intervention tailored to a specific person. Specifically, *CausalGait* has 3 key components: 1) *Global Causal Graph*, which encodes standard clinical rules into a symbolic Directed Acyclic Graph (DAG); 2) *Patient-Specific Graph*, which translates measured gait data into Patient-Specific Graphs via a node activation mechanism; and 3) *Causal Inference and Interpretation*, which performs conditional probability estimation and constrained Bayesian analysis for personalized annotation. We evaluated *CausalGait* on a dataset of 200 children with spastic CP collected at a credentialed gait lab. Results show that *CausalGait* outperforms the Data-Driven baseline on all three metrics: achieving high faithfulness (0.868), low infidelity (0.0023), and strong perturbation-based ranking consistency (normalized AOPC = 0.950). A case study further demonstrates the

interpretation capability of *CausalGait* through medical expert analysis.

CCS Concepts

• **Human-centered computing** → *Ubiquitous and mobile computing*; • **Applied computing** → **Health informatics**; • **Computing methodologies** → **Causal reasoning and diagnostics**.

Keywords

Knowledge Graph, Causal Analysis, Bayesian Network, Personalized Interpretation

1 Introduction

Clinical gait interpretation of human motion data is important because it can support various downstream analytical tasks, from abnormality detection in both clinical and real-world environments [6] to strategic intervention planning for neuromuscular disorders such as Cerebral Palsy (CP) [9] and Muscular Dystrophy, thereby evolving toward precision medicine tailored to individual patients.

Realizing such precision medicine requires precise identification of the contributing neuromuscular impairments for each observed gait abnormality [3]. The four primary neuromuscular impairments in spastic CP are: muscle weakness, short muscle-tendon unit, spasticity, and impaired selective motor control [3, 4]. While observational and 3D gait analysis can successfully identify gait abnormalities, evidence-based decision-making and treatment planning require understanding *why* they occur [9].

Prior research has made attempts to model the relationship between neuromuscular impairments and gait abnormalities



This work is licensed under a Creative Commons Attribution 4.0 International License.

HumanSys '26, Austin, TX, USA

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2959-1/26/10

<https://doi.org/10.1145/3842436.3843806>

[10]. However, these models are limited to either fixed clinical rules, such as the Clinical GENI tool [3], or population-based data-driven models that link gait observations with the underlying neuromuscular impairments [10], which can lead to one-size-fits-all intervention strategies that have limited interpretability towards a specific person. Neither method alone can provide the personalized, interpretable analysis to identify contributing neuromuscular impairments required for strategic intervention.

To close this gap and explore the possibility of providing personalized clinical insights, we introduce *CausalGait*, a new framework that integrates clinical rules and measured gait data into personalized causal knowledge graphs, aiming to assist medical experts with more strategic intervention tailored to a specific person.

We summarize our contributions as follows:

- We develop *CausalGait*, a pioneering framework that embeds clinical rules into population-level *Global Causal Graph* (GCG) and transforms personalized gait data into *Patient-Specific Graphs* (PSG) to provide gait interpretation tailored to a person's unique gait and neuromuscular impairments.
- We introduce *Patient-Specific Graph Instantiation* that models the relationship between patient-specific impairment and gait abnormalities through a node activation mechanism to provide causality-informed gait interpretation, and identify isolated nodes for potential future discovery of new causal relationships.
- We evaluate the explanation quality of *CausalGait* using three quantitative metrics, and a comparison against a Data-Driven baseline, and further demonstrate its feasibility for clinical decision support through a representative case study.

2 CausalGait System Design

The end-to-end framework design of *CausalGait* is shown in Fig. 1. *CausalGait* consists of three components. First, GCG encodes the impairment-abnormality relationships specified by Clinical GENI [3, 4] as a directed acyclic graph, which serves as a shared template for all patients. Second, we introduce a node activation mechanism that instantiates the GCG into PSGs for each patient, which retains only the impairment and gait abnormalities observed in that patient, so that subsequent inference is restricted to the relationships actually present in this patient. An abnormality observed without any of its GENI-specified parent impairments is retained as an orphan abnormality node, flagging cases that current clinical rules cannot explain. Third, *Causal Inference & Interpretation* computes conditional probabilities for population-level insights and applies BN to PSGs for personalized gait interpretation informed by neuromuscular impairments.

2.1 Global Causal Graph

To formalize the relationship between neuromuscular impairments and gait abnormalities from a global perspective, we construct the GCG that transforms qualitative clinical rules into a symbolic DAG based on the Clinical GENI Tool [3, 4], as shown on the left side of Fig. 2.

Specifically, GCG can be formalized as $G_E = (V_I, V_A, E)$, where V_I represents the set of impairment nodes, V_A represents the set of abnormality nodes, and E represents the set of directed causal edges. Each edge represents an expert-supported causal relationship from a specific neuromuscular impairment to a corresponding gait abnormality, which can be formulated as $(i, a) \in E \leftrightarrow \text{GENI}(i, a) = 1$, where $i \in V_I$ and $a \in V_A$.

2.2 Patient-Specific Graph

To better provide personalized clinical insights, *CausalGait* instantiates PSG by performing a node activation mechanism based on the GCG. A concrete example is shown on the right side of Fig. 2.

Since each patient only contains a subset of the impairment-abnormality relationships, PSG only preserves the causal edges among the activated nodes. Formally, we define $G_p = (V_p, E_p)$, where $G_p \subseteq G_E$, and $V_p = \{v \in V_I \cup V_A \mid x_v^p = 1\}$ represents the activated nodes.

For each patient p , *CausalGait* assigns a binary activation state to every node as x_v^p : if node v is activated for patient p , then $x_v^p = 1$. Otherwise, 0. Here, v can represent either an impairment or an abnormality node. The activation of each node is performed independently. In this case, PSG can preserve instances when a gait abnormality is present without related neuromuscular impairments. This point is particularly important because such an inconsistency may indicate clinically meaningful corner cases and explore possibilities for clinical rule refinement. Specifically, we define the isolated gait abnormality nodes as *Orphan Abnormality Nodes* (OAN). For an OAN a , we denote its parent impairment node set in the PSG as: $\text{Parent}(a) = \{i \mid i \rightarrow a \in G_p\}$.

For patient p , if an activated gait abnormality node a is observed in the patient, but none of its parent neuromuscular impairment nodes are activated, then a is defined as an orphan abnormality: $\text{Orphan}(a, p) = 1 \leftrightarrow x_a^p = 1$ and $\sum_{i \in \text{Parent}(a)} x_i^p = 0$.

In this context, orphan gait abnormalities may reveal previously unexplored mechanisms of pathological impairment, thereby facilitating potential causal relationship discovery in the future.

2.3 Causal Inference and Interpretation

To support better clinical interpretation of the causal relationship at both the global and the patient instance level, *CausalGait* quantifies the strength of each causal edge at

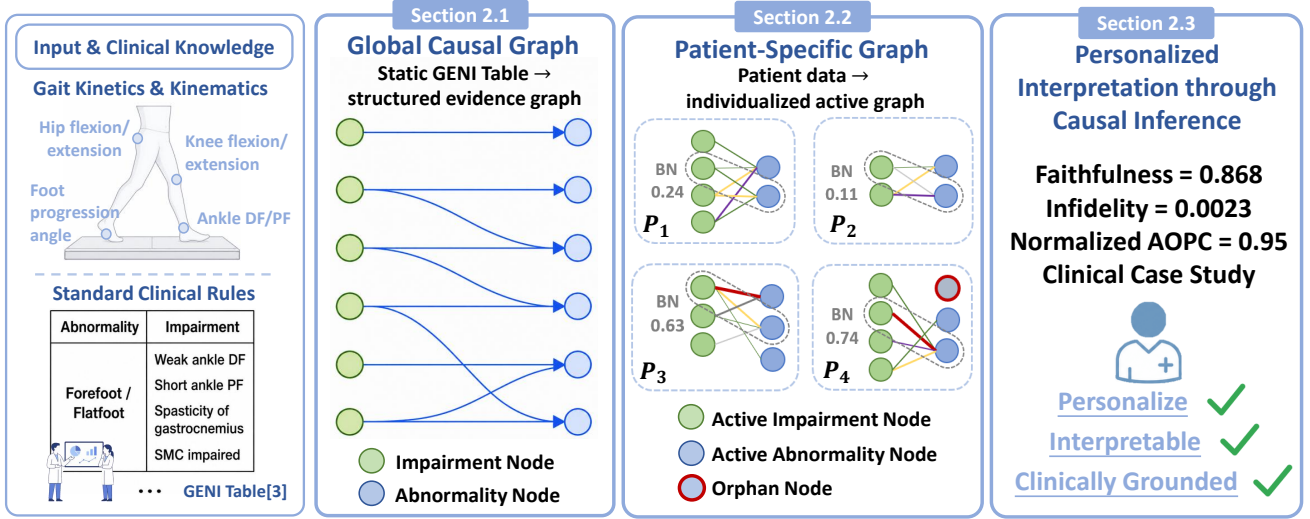


Figure 1: System workflow of *CausalGait*. *CausalGait* takes the clinical rules and gait kinematics/kinetics data as input, generates the GCG and PSG, and performs causal analysis from different perspectives.

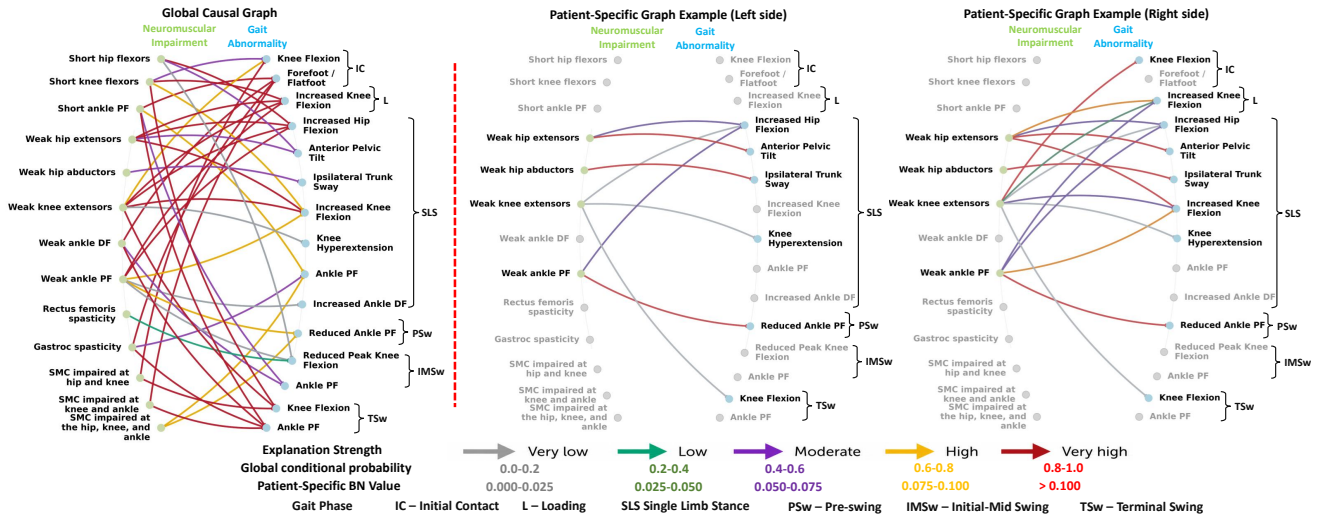


Figure 2: GCG and PSG examples. The GCG captures population-level impairment-abnormality relationships, while PSGs instantiate personalized active relations for the left and right sides. Edge color denotes conditional probability in GCG, and model-estimated BN value in PSG. For example, Weak ankle plantarflexors (PF) can alter ankle dorsiflexion (DF) control during the Single Limb Stance phase, which may propagate to abnormal knee and hip flexion when the limb is loaded, explaining why a distal impairment can contribute to multiple gait abnormalities.

both GCG and PSGs, and further identifies misalignment patterns.

2.3.1 Global statistical analysis using conditional probability estimation. To provide population-based statistical insights for global interpretation, *CausalGait* calculates how frequently abnormality a appears among patients with impairment i in GCG. This global conditional probability can

prioritize the relationship for a specific kind of neuromuscular impairment and abnormality, further defined as:

$$\hat{P}(a | i) = \frac{\sum_{p=1}^N \mathbf{1}(x_i^p = 1 \text{ and } x_a^p = 1)}{\sum_{p=1}^N \mathbf{1}(x_i^p = 1)} \quad (1)$$

The numerator of Equation 1 represents the number of patients who have both impairment i and abnormality a , while the denominator represents the number of patients who have

impairment i . Therefore, $\hat{P}(a | i)$ measures the population-level occurrence rate of abnormality a under the condition that impairment i is observed.

CausalGait uses this probability to quantify the empirical strength of each causal relationship. For example, if two different neuromuscular impairments are both associated with the same gait abnormality, comparing their global conditional probabilities allows *CausalGait* to determine which impairment provides a stronger population-level explanation for that abnormality.

2.3.2 Personalized interpretation using causality-informed Bayesian Networks. To support patient-specific analysis and prioritize the most likely root causes for a given gait abnormality, *CausalGait* uses BN analysis to provide a weighted interpretation for the directed edges in each PSG, so that it can show not only whether a potential clinical causal relationship exists, but also how strongly this relationship explains the abnormality under the current patient context.

For an abnormality node a , *CausalGait* first extracts all of its parent impairment nodes from the GCG, denoted as $\text{Parent}(a)$.

For each patient p , *CausalGait* observes which impairment nodes in the PSG are activated, thereby forming a parent activation pattern. For example, if abnormality a has three parent impairments, then $\mathbf{z} = [1, 0, 1]$ indicates that only the first and third parent impairments are activated.

Because all impairments and abnormalities relationship are side-specific, the left and right limbs of each participant are treated as separate patient-side observations in the probability estimation.

Based on all PSGs, *CausalGait* estimates the pattern conditioned probabilities using constrained Bayesian estimation with positive monotonicity constraints [11]. Let $\hat{P}_a(\mathbf{z})$ denote the estimated probability of abnormality a under parent pattern $\mathbf{z}$. For an activated edge $i \rightarrow a$, its patient-specific BN value is

$$\text{BN}_{i \rightarrow a}^{(p)} = \hat{P}_a(\mathbf{z}_p) - \hat{P}_a(\mathbf{z}_p^{(-i)}) \geq 0,$$

where $\mathbf{z}_p^{(-i)}$ is obtained by setting impairment i to zero while keeping all other parent states unchanged. The BN value measures the corresponding decrease in the estimated abnormality probability. A larger BN value indicates a stronger contribution of impairment i under current patient context.

2.3.3 Potential causal relationship discovery based on Orphan Rate. While global statistical analysis (Section 2.3.1) and personalized interpretation (Section 2.3.2) provide quantitative information at global and patient-specific levels, respectively, *CausalGait* further uses *Orphan Rate* (OR) to quantify the proportion of observed abnormalities that are not explained by activated parent impairments. This measure helps clinicians

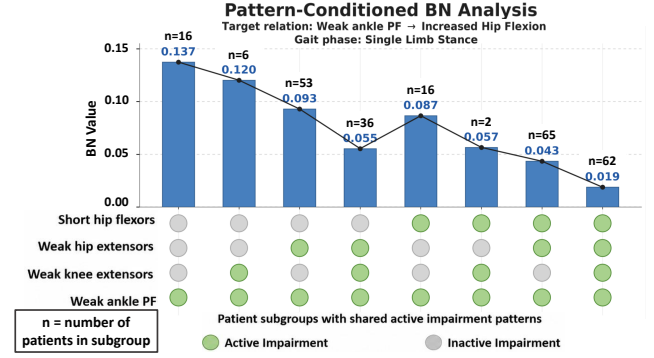


Figure 3: Pattern-stratified BN analysis for the target relation Weak ankle PF $\rightarrow$ Increased Hip Flexion. Each column represents a patient subgroup with a shared impairment pattern. The matrix shows active/inactive impairments, bar height, and the overlaid line indicates the BN value. n denotes subgroup size.

identify corner cases that may warrant further investigation and refinement of the current clinical rules.

For an abnormality a , let P_a denote the set of patient instances in which abnormality a is observed, i.e., $P_a = \{p \mid x_a^{(p)} = 1\}$. The OR of abnormality a is defined as:

$$\text{OR}(a) = \frac{\left| \left\{ p \in P_a \mid \sum_{i \in \text{Parent}(a)} x_i^{(p)} = 0 \right\} \right|}{|P_a|} \quad (2)$$

For abnormality a , this metric measures the proportion of observed instances that cannot be explained by any activated parent impairment. A high OR suggests that this abnormality often appears without any activated impairment-based explanation. Finally, *CausalGait* ranks abnormalities according to their ORs to provide clinicians with an opportunity to explore potential causal relationships.

3 CausalGait Evaluation

To comprehensively evaluate the *CausalGait* framework and explore the personalized interpretation capability, we construct PSGs from a credentialed 3D gait analysis lab dataset of 200 spastic CP patients who meet inclusion criteria, perform causal inference and analysis, quantitatively evaluate the patient-specific BN values, and verify their clinical interpretation through a representative case study. This study was reviewed and approved by the Institutional Review Board (IRB) of [Stanford University] under protocol number [11910]. All participants provided informed consent before participation.

3.1 Experimental Setup

3.1.1 Inclusion Criteria. Inclusion criteria is a diagnosis of spastic CP, age 3-17 y/o, ability to walk independently (i.e., without an assistive device or external assist) for at least 6

meters, and having a physical exam of neuromuscular impairments including manual muscle test (MMT), passive range of motion (PROM), the modified Tardieu Scale (MTS), and the Selective Control Assessment of the Lower Extremity (SCALE) [5]. Exclusion criteria include a history of surgical intervention, including lower extremity (LE) orthopedic surgery or selective dorsal rhizotomy (SDR) within 1 year of gait analysis and/or botulinum toxin type A (BoNT-A) injection within 1-year of gait analysis, and no use of baclofen at the time of testing.

3.1.2 Comparison Baseline. We use Random Forest (RF) with local SHAP explanations as the Data-Driven baseline. A separate RF is trained for each gait abnormality using binary impairment indicators as input. Each model contains 300 trees and uses balanced class weights.

Let K_a denote the number of impairments associated with abnormality a in Clinical GENI criteria. For each patient, the Data-Driven baseline ranks impairments by their positive SHAP values and retains the top K_a , so that the number of explanations matches the Clinical GENI criteria.

3.1.3 Evaluation Metrics. We evaluate *CausalGait* and the Data-Driven baseline using three metrics from existing studies: Faithfulness [2], Infidelity [12], and normalized Area Over the Perturbation Curve (AOPC) [8]. Each metric is defined below.

Faithfulness measures the correlation between the explanation scores and the changes in the estimated abnormality probability when impairments are removed [2]. For abnormality cases of each patient, let S be a pair of activated parent impairments, and define

$$A(S) = \sum_{i \in S} s(p, i, a), \quad D(S) = f_{p,a}(\emptyset) - f_{p,a}(S),$$

where $f_{p,a}(S)$ is the estimated abnormality probability after the impairments in S are removed.

$$\text{Faithfulness} = \text{mean}_{(p,a)} [\rho_{|S|=2}(A(S), D(S))].$$

where ρ denotes the Pearson correlation coefficient. We use impairment pairs because each score in *CausalGait* is already defined using a single impairment perturbation. The same pairwise evaluation is applied to the Data-Driven baseline. A higher Faithfulness indicates stronger agreement between the explanation scores and the observed probability changes.

Infidelity measures the squared error between the summed explanation scores and the decrease in the estimated abnormality probability when the corresponding impairments are jointly removed [12]:

$$\text{Infidelity} = \text{mean}_{(p,a)} \left[\text{mean}_{|S|=2} (A(S) - D(S))^2 \right].$$

Table 1: Comparison of explanation consistency between *CausalGait* and Data-Driven baseline.

Metric	<i>CausalGait</i>	Data-Driven baseline
Faithfulness	0.868	0.519
Infidelity	0.0023	0.0164
Normalized AOPC	0.950	0.777

A lower Infidelity indicates that the perturbation-weighted explanation scores more closely match the resulting probability changes.

Normalized AOPC tests whether the ranking based on the original explanation scores remains effective when multiple impairment indicators are perturbed cumulatively [8]. Let $\pi = (\pi_1, \dots, \pi_m)$ denote the selected impairments ranked by decreasing raw explanation score. Here, m is determined by the number of impairments associated with abnormality a in Clinical GENI. For each patient-abnormality case (p, a) , we compute

$$\text{AOPC}_{p,a}(\pi) = \frac{1}{m+1} \sum_{k=0}^m [f_{p,a}(\emptyset) - f_{p,a}(\{\pi_1, \dots, \pi_k\})],$$

where $\{\pi_1, \dots, \pi_0\} = \emptyset$. We then normalize this value by the exact minimum and maximum AOPC attainable over all $m!$ removal orders:

$$\text{nAOPC}_{p,a} = \frac{\text{AOPC}_{p,a}(\pi) - \text{AOPC}_{p,a}^{\min}}{\text{AOPC}_{p,a}^{\max} - \text{AOPC}_{p,a}^{\min}}.$$

The normalized AOPC is the mean of $\text{nAOPC}_{p,a}$ across all patient-abnormality cases. A higher normalized AOPC indicates that a method's ranking is closer to the optimal removal order under its own fitted model.

3.2 Quantitative Evaluation Results

Table 1 presents the quantitative results. *CausalGait* achieves a Faithfulness of 0.868 and an Infidelity of 0.0023. Across patients and observed abnormalities, impairment pairs with larger summed BN values tends to correspond to larger probability decreases, while the summed values remain close to the changes produced by setting both impairment states to 0. The normalized AOPC of 0.950 shows that the ranking generally places impairments with greater estimated influence earlier and approaches the best removal order for each case. Since these metrics are calculated using probabilities from the fitted model, they measure the internal consistency of the explanation scores rather than clinical accuracy.

In a future study, clinical experts will evaluate each impairment–abnormality edge for a randomly selected group of patients. Their assessments will serve as an expert-derived reference standard for evaluating the accuracy of *CausalGait*.

3.3 Case Study

To demonstrate how *CausalGait* translates population-level GCG into patient-specific interpretation and prioritizes competing impairment explanations, we analyze a representative patient in collaboration with a medical expert using the example shown in Fig. 2.

GCG Interpretation. In the spastic CP population, muscle weakness is one of the most commonly identified neuromuscular impairments [3]. Ankle Dorsiflexor Weakness reduces Ankle Dorsiflexion in swing, causing the ankle to remain plantarflexed at Initial Contact and resulting in Flatfoot or Forefoot Contact. Absence of Heel Contact eliminates the Heel Rocker, negatively impacting forward gait progression and energy expenditure [1, 3, 7].

PSG Interpretation. In this specific patient, Weak Ankle DF is present on the left limb but absent on the right, highlighting the importance of bilateral assessment. Short Ankle PF is present bilaterally, further limiting the Ankle's ability to achieve neutral Dorsiflexion in Swing and at Initial Contact and contributing to Forefoot/Flatfoot IC. As the knee progresses toward extension in late swing, increased stretch on the gastrocnemius may further limit neutral DF and Heel Contact. Knee Hyperextension in Single-Limb Stance was associated with Weak Knee Extensors. While Weak Knee Extensors had a BN of 0.022, the frequency of this abnormality in the sample studied was 5.3%, therefore indicating that it is not as common a gait abnormality as Increased Knee Flexion in Single-Limb Stance, which occurred in 81.1%. Muscle Weakness was tested using MMT. While MMT is commonly used clinically due to its simplicity and lack of required equipment, it has limitations in detecting strength impairments in large muscle groups such as the quadriceps. More quantitative methods such as dynamometry or gold standard isokinetic testing would be more sensitive for detecting muscle weakness of the quadriceps, though these are not always clinically available due to equipment costs. It is possible that MMT underestimated the magnitude of knee extensor weakness in this sample, and future studies should compare these relationships using dynamometry or isokinetic testing.

4 Conclusion

This paper introduces *CausalGait*, a novel framework for personalized and causality-informed gait interpretation. *CausalGait* addresses the interpretation gap between rule-based clinical reasoning and data-driven models through causality-informed knowledge graphs that integrate both clinical rules and gait data. For each patient, the global graph is instantiated using the active impairments, producing patient-specific BN values and impairment rankings. Evaluation reveals that *CausalGait* outperformed the Data-Driven baseline across all three metrics, showing stronger consistency between its

explanation scores and model behavior. In the future, we plan to further develop a multi-agent collaboration system to analyze the implicit pathological features embedded in *CausalGait* and generate personalized treatment recommendation reports.

References

- [1] M. Attias, A. Bonnefoy-Mazure, G. De Coulon, L. Cheze, and S. Armand. 2023. Toe-walking and its impact on first and second rocker in gait patterns with different degrees of artificially emulated soleus and gastrocnemius contracture. *Gait & Posture* 105 (2023), 104–109. doi:10.1016/j.gaitpost.2023.07.285
- [2] Umang Bhatt, Adrian Weller, and José M. F. Moura. 2020. Evaluating and aggregating feature-based model explanations. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, California, 3016–3022.
- [3] Kylie Clewes, Yiwen Dong, Mary Meyer, Evan Lowe, Kornél Schadt, and Jessica Rose. 2025. Clinical Gait Evaluation with Neuromuscular Impairments (Clinical GENI) for spastic cerebral palsy. *Frontiers in Human Neuroscience* Volume 19 - 2025 (2025). doi:10.3389/fnhum.2025.1637164
- [4] Kylie Clewes, Claire Hammond, Yiwen Dong, Mary Meyer, Evan Lowe, and Jessica Rose. 2024. Neuromuscular impairments of cerebral palsy: contributions to gait abnormalities and implications for treatment. *Frontiers in Human Neuroscience* Volume 18 - 2024 (2024). doi:10.3389/fnhum.2024.1445793
- [5] Eileen G Fowler, Loretta A Staudt, Marcia B Greenberg, and William L Oppenheim. 2009. Selective Control Assessment of the Lower Extremity (SCALE): development, validation, and interrater reliability of a clinical tool for patients with cerebral palsy. *Developmental Medicine & Child Neurology* 51, 8 (2009), 607–614.
- [6] Janire Otamendi, Asier Zubizarreta, and Eva Portillo. 2023. Machine learning-based gait anomaly detection using a sensorized tip: an individualized approach. *Neural computing and applications* 35, 24 (2023), 17443–17459.
- [7] Firooz Salami, Sarah Campos, Arik Rehani Musagara, and Sebastian I. Wolf. 2024. Modeling foot rockers via functional calibration for use in clinical gait analysis. *Gait & Posture* 111 (2024), 122–125. doi:10.1016/j.gaitpost.2024.04.022
- [8] Wojciech Samek, Alexander Binder, Gregoire Montavon, Sebastian Lapuschkin, and Klaus-Robert Müller. 2017. Evaluating the visualization of what a deep neural network has learned. *IEEE Transactions on Neural Networks and Learning Systems* 28, 11 (11 2017), 2660–2673.
- [9] Michael H Schwartz and Andrew G Georgiadis. 2025. Evidence Based Gait Analysis Interpretation Tools (EB-GAIT) treatment recommendation and outcome prediction models to support decision-making based on clinical gait analysis data. *PloS one* 20, 7 (2025), e0328036.
- [10] Michael H. Schwartz, Katherine M. Steele, Andrew J. Ries, Andrew G. Georgiadis, and Bruce A. MacWilliams. 2022. A model for understanding the causes and consequences of walking impairments. *PLOS ONE* 17, 12 (12 2022), 1–29. doi:10.1371/journal.pone.0270731
- [11] Gao Xiaoguang, Yang Yu, and Guo Zhigao. 2019. Learning Bayesian networks by constrained Bayesian estimation. *Journal of Systems Engineering and Electronics* 30, 3 (2019), 511–524. doi:10.21629/JSEE.2019.03.09
- [12] Chih-Kuan Yeh, Cheng-Yu Hsieh, Arun Sai Suggala, David I. Inouye, and Pradeep Ravikumar. 2019. On the (In)fidelity and Sensitivity for Explanations. arXiv:1901.09392 [cs.LG] <https://arxiv.org/abs/1901.09392>