

# VIGMA: An Open-Access Framework for Visual Gait and Motion Analytics

Kazi Shahrukh Omar<sup>✉</sup>, Shuaijie Wang, Ridhuparan Kungumaraju, Tanvi Bhatt, Fabio Miranda<sup>✉</sup>

**Abstract**—Gait disorders are commonly observed in older adults, who frequently experience various issues related to walking. Additionally, researchers and clinicians extensively investigate mobility related to gait in typically and atypically developing children, athletes, and individuals with orthopedic and neurological disorders. Effective gait analysis enables the understanding of the causal mechanisms of mobility and balance control of patients, the development of tailored treatment plans to improve mobility, the reduction of fall risk, and the tracking of rehabilitation progress. However, analyzing gait data is a complex task due to the multivariate nature of the data, the large volume of information to be interpreted, and the technical skills required. Existing tools for gait analysis are often limited to specific patient groups (e.g., cerebral palsy), only handle a specific subset of tasks in the entire workflow, and are not openly accessible. To address these shortcomings, we conducted a requirements assessment with gait practitioners (e.g., researchers, clinicians) via surveys and identified key components of the workflow, including (1) data processing and (2) data analysis and visualization. Based on the findings, we designed VIGMA, an open-access visual analytics framework integrated with computational notebooks and a Python library, to meet the identified requirements. Notably, the framework supports analytical capabilities for assessing disease progression and for comparing multiple patient groups. We validated the framework through usage scenarios with experts specializing in gait and mobility rehabilitation. VIGMA is available at [github.com/komar41/VIGMA](https://github.com/komar41/VIGMA).

**Index Terms**—Gait, stroke, rehabilitation, visual analytics, data processing, multivariate data, ensemble, survey, computational notebook.

## I. INTRODUCTION

Gait disorders are prevalent in older adults, with studies revealing that 82% of people aged 85 years or older suffer from some sort of gait anomaly [1], [2]. Conditions such as Parkinson's [3] or stroke [4] significantly impact patients' gait, causing notable deviations from typical motor movements such as shuffling or asymmetric gait. Additionally, gait impairments are increasingly observed in children with cerebral palsy [5] and spina bifida [6]. The analysis of gait presents a complex challenge due to its variability within and across different patient groups, coupled with the diverse types of data collected for each individual. Consequently, tools that straddle the entire gait data lifecycle (processing, analyzing, and visualizing) can play a pivotal role in understanding the biomechanics of a patient, monitoring disease progression, and implementing appropriate rehabilitation strategies.

Kazi Shahrukh Omar, Ridhuparan Kungumaraju, and Fabio Miranda are with the Department of Computer Science, and Shuaijie Wang and Tanvi Bhatt are with the Department of Applied Health Sciences, University of Illinois Chicago, IL 60607, USA (emails: komar3@uic.edu; sjwang4@uic.edu; rkungu2@uic.edu; tbhatt6@uic.edu; fabiom@uic.edu;)

Gait data is collected via a diverse combination of hardware (e.g., infrared (IR) cameras, force plates) and software tools (e.g., Visual3D [7], Vicon Nexus [8]) that generate a wide variety of time-series data, including kinetics (ground reaction forces, joint moment), kinematics (motion, joint angles), and electromyography, as well as single-valued spatiotemporal parameters (e.g., step length, swing time). Each of these data types plays a crucial role in various tasks associated with the analysis of patients' gait. For instance, kinetic measures and joint moments can provide insights into several musculoskeletal diseases [9], whereas kinematic measures are particularly useful for assessing fall risk in elderly individuals [10]. Gait data also includes raw trial videos, which, when analyzed alongside other time-series data, provide additional context to interpret anomalies or validate findings. This diversity of data underscores the need for comprehensive tools that encompass all of these data types to enable a more detailed analysis of patients' gait.

Therefore, analyzing gait data constitutes a multifaceted challenge, as the workflow involves several data processing, and data analysis and visualization tasks. Data processing tasks involve format harmonization, imputing missing values, filtering noise, whereas analysis and visualization tasks include analyzing statistical measures, comparing patient groups, finding anomalies, and analyzing disease progression. Both sets of tasks are vital in the overall workflow of practitioners. However, existing tools and techniques predominantly support only a subset of tasks and lack important analysis and visualization capabilities, such as comparing patient groups and analyzing disease progression. Furthermore, the lack of open access to tools utilized by practitioners in different labs hinders collaborative efforts between groups.

To tackle these challenges, this paper presents the open-access Visual Gait and Motion Analytics (VIGMA) framework. VIGMA was designed in close collaboration with domain experts and is the result of a survey with over 20 gait practitioners, including clinicians, researchers, administrators, data analysts, faculty members, and students. The survey allowed us to surface detailed workflows for analyzing gait data that involved a wide variety of tasks, including data processing, and data analysis and visualization. In response to the wide needs of gait practitioners, VIGMA is designed as both (a) a library for use in computational notebooks, enabling data processing tasks, and (b) a visual analytics system for data analysis and visualization. The visual analytics system supports the visualization of multivariate time-series ensemble gait data, along with the analysis of video data and spatiotemporal parameters with tailored interaction techniques

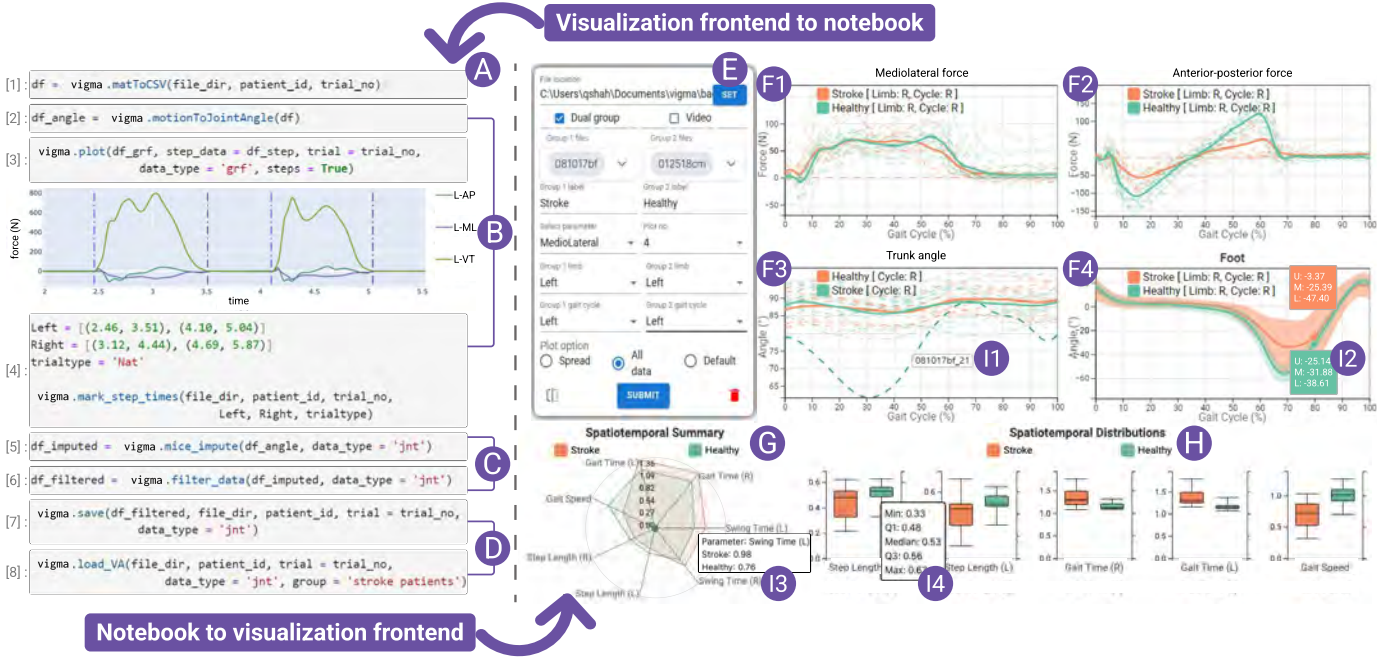


Fig. 1. **VIGMA's** components. Using Jupyter Notebook and VIGMA's library, users begin by **A** harmonizing collected data to standard CSV format. Afterwards, they perform essential **B** feature extraction steps (e.g., extracting joint angles from motion data, extracting step times from ground reaction forces). Next, they prepare the data by performing tasks like **C** imputing missing values and filtering out noise. Finally, **D** users save the processed data and upload it to the visual analytics system, categorized by patient group. The visual analytics system consists of **E** a control panel that lets users select ensemble data files and select chart configurations. The system is composed of four time-series ensemble views (**F1** - **F4**) to support the analysis of time-series gait data (e.g., joint angles). Additionally, it features **G** a spatiotemporal summary view and **H** a spatiotemporal distribution view to support single-valued gait parameter analysis. Users can choose to alternate between the spatiotemporal views (**G**, **H**) and the video exploration view (Fig. 8) using the checkbox from the control panel **E** and get access to raw trial videos. **I1** to **I4** highlight interactive details-on-demand features of the system. Users can identify anomalies (e.g., **I1**), trace errors to patients' trial videos, and return to the computational notebook to correct the data.

to support tasks such as group comparison and tracking disease progression. Fig. 1 illustrates the VIGMA framework interface, highlighting its individual components and their functionalities. This paper also presents an evaluation of VIGMA that includes three usage scenarios created in collaboration with domain experts, as well as experts' quantitative and qualitative feedback on the framework. Our contributions can be summarized as follows:

- We present a comprehensive survey yielding requirements for gait data processing, analysis, and visualization.
- We contribute a Python library that provides functionalities to satisfy gait data processing requirements.
- We contribute a novel visual analytics system that supports the data analysis and visualization requirements.
- We present three usage scenarios created in collaboration with domain experts to evaluate the efficacy of the system.

This paper is organized as follows: Section II presents background on gait data and analysis; Section III reviews related work; Section IV presents the result of our survey and requirements; Section V details the VIGMA framework; Section VI presents the evaluation; Section VII concludes the paper.

## II. BACKGROUND

Gait, or walking, is a vital social activity controlled by an intricate combination of brain activities and limb movements.

It is commonly studied in the elderly, children, and athletes due to the unique challenges and needs these groups present. Practitioners collect gait data, categorized into videos, kinematics, kinetics, electromyography, and spatiotemporal parameters. Monthly intervals are typical for data collection, varying based on disease and severity [11], [12].

Practitioners tailor their collection of data types to the specific needs of different patient groups. For instance, post-stroke patients exhibit a spectrum of characteristics, such as decreased walking speed and shorter stride length [17]. They also encounter heightened fall risks due to impaired balance control and muscle weaknesses [18]. In the case of cerebral palsy, patients showcase distinctive walking patterns (e.g., jump knee gait, equinus [19]), which can be identified by analyzing kinematic gait data [20]. Parkinson's patients, on the other hand, face noticeable gait deterioration in kinematic (e.g., ankle, knee angle) and spatiotemporal parameters (e.g., reduced step length and arm swing) [21].

Following such diversity of needs, data is collected in different scenarios, including regular and perturbed (slip/trip) walking. Fig. 2 illustrates a typical data collection process [13]–[16]. It involves a walkway with embedded movable platforms, which rest on force plates to record kinetic data. These platforms unlock once participants step on them during slip trials, and participants are protected by a safety harness during the trials. An infrared camera-based motion capture system records video and kinematic data by tracking markers placed

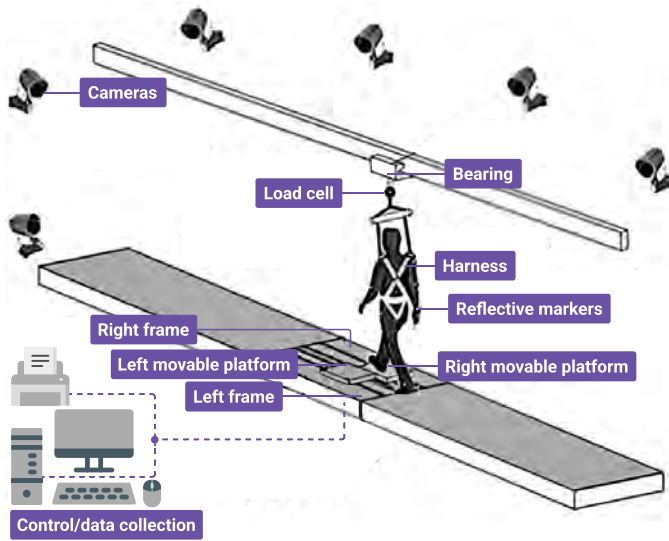


Fig. 2. Data collection process [13]–[16]. A 7-meter walkway with embedded low-friction movable platforms is used to collect slip and walking data. The platforms move along low-friction aluminum tracks beneath the surface and rest on four force plates. The force plates record ground reaction forces, which also trigger the platforms to create the slipping effect. The platforms lock firmly during regular walking and unlock electromechanically without the participants’ awareness in the slip trial. Participants are protected by a safety harness connected to a low-friction trolley-and-beam system above the walkway. Kinematic data from all trials is recorded by an eight-camera motion capture system, synchronized with the force plates and load cell data.

on the body. Electromyography (EMG) data is collected using electrodes placed on the skin over the muscles of interest, which detect electrical activity during walking. Spatiotemporal data, which includes spatial parameters (e.g., step length) and temporal parameters (e.g., cadence), is often derived from kinematic or kinetic data. All these different data types are synchronized to provide a comprehensive analysis of gait.

The collected gait data types fall into the categories of time-series or single-valued, with spatiotemporal parameters belonging to the latter. To process and analyze these two different types of data, different software systems are used, facilitating processing, analysis, and visualization of the collected data. In terms of analysis, practitioners typically focus on tasks such as: (1) comparing patients’ gait characteristics with those of healthy control groups or similar patients to identify deviations and abnormalities [22]–[24], (2) assessing gait changes over time to monitor disease progression and evaluate the effectiveness of rehabilitation strategies [25]–[27], (3) statistical analysis [28], [29], and (4) finding anomalies [29], [30].

### III. RELATED WORK

In this section, we review related work across four categories. First, we discuss the existing works on visualization for gait; next, we explore literature for visualization for time-series data; then, we investigate open-access and commercial tools that support gait data processing, analysis, and visualization workflows; and finally, we review existing research on computational notebooks to support data processing tasks and how they are combined with visualization interfaces.

**Visualization for gait data.** Gait data can be grouped into (1) time-series [31] and (2) single-valued parameters [32].

Spatiotemporal data are considered single-valued parameters, while kinetic, kinematic, and electromyography data are classified as time-series. Existing approaches commonly use line charts [31], [33]–[35] for time-series gait data, radar charts [32], [36] for single-valued parameters, and box plots [32], [33] or parallel coordinate plots [37] for the distribution of single-valued parameters. However, these approaches often lack the interactivity and human-in-the-loop capabilities provided by traditional visual analytics tools. In contrast, tools like MotionExplorer [38], MotionFlow [39], and Chen et al. [40] support interactivity to provide insight into patient’s gait and help researchers cluster similar human poses. However, these tools do not provide any insight into quantitative gait measures (e.g., kinetic, kinematic, spatiotemporal) that practitioners use for the analysis and treatment of patients.

Existing literature introduced several novel visual analytics tools for quantitative gait analysis. The NE-Motion visual analytics tool [41] is designed to compare the gait of healthy individuals with that of stroke patients experiencing upper limb impairments. KAVAGait [42] supports the analysis of kinematic (e.g., ground reaction forces) and spatiotemporal gait in various patient groups. GaitViewer [43] proposes a structured approach to storing numerical gait data, facilitating collaboration between research labs, and providing an interactive visual interface for spatiotemporal gait parameters. gaitXplorer [44] is a visual analytics tool supporting kinematic gait analysis of cerebral palsy patients. It supports classifying patients into different groups (e.g., jump knee, equinus) and provides insights into the specific data regions that contributed to the classification. Although useful, these tools offer very limited data processing capabilities. None provides comprehensive multivariate feature support, and only KAVAGait and gaitXplorer support hierarchical data management. Also, only GaitViewer supports feature extraction, but is limited to a few spatiotemporal parameters. Furthermore, these tools lack important data analysis and visualization capabilities. Specifically, KAVAGait and NE-Motion do not support disease progression analysis; GaitViewer lacks patient group comparison, while gaitXplorer lacks both. Lastly, none of these tools are openly available.

**Visualization for time-series data.** Time-series gait is multivariate, including diverse attributes such as joint angles, forces, and electromyography signals, analyzed across different limbs and normalized over gait cycles. These attributes are often grouped into ensembles for analyzing similar patient groups (e.g., stroke/healthy) collectively or comparing between two groups. Numerous studies have delved into the visual analysis of multivariate time-series data [45]–[51], as well as time-series ensemble data [52]–[54]. In gait studies, line charts are the most widely used visualization technique for time-series data [7], [42], [44], making them a familiar choice for gait practitioners in research labs or clinical settings. In this work, we explore design variations and interactions with line charts to effectively represent multivariate and time-series ensemble gait data, supporting the analysis of individual patient’s gait as well as tasks such as disease progression and group comparisons.

Time-series gait analysis also includes videos that offer raw



TABLE I

OVERVIEW OF HOW OPEN-ACCESS AND COMMERCIAL TOOLS, ALONG WITH TOOLS FROM EXISTING LITERATURE, ADDRESS DIFFERENT DATA PROCESSING, ANALYSIS, AND VISUALIZATION FEATURES. OPEN SOURCE REFERS TO SOFTWARE THAT IS FREELY ACCESSIBLE AND ALLOWS CODE INSPECTION. THE NOTATIONS USED IN THE TABLE SIGNIFY THE FOLLOWING: ✓ - FEATURE IS FULLY SUPPORTED, △ - FEATURE IS PARTIALLY SUPPORTED, AND EMPTY CELLS INDICATE THAT THE RESPECTIVE TOOL DOES NOT SUPPORT THE FEATURE.

Features	Kinovea [62]	bioMechZoo [61]	OpenCap [63]	GaitLab [64]	Vicon Nexus [8]	TEMPLO [65]	BioSensics [66]
Format harmonization		△					
Feature extraction	△	✓	✓	△	✓	△	△
Missing value imputation		✓			✓		
Noise filtering		✓			✓		
Gait cycle normalization		✓				✓	
Hierarchical data management					✓		
Multivariate feature support	△	✓	✓	△	✓	✓	△
Group comparison		△					
Disease progression		△					
Statistical analysis		△					
Anomaly analysis		△					
Access to raw video data	✓		✓			✓	
Interactive analysis	△	△				△	
Open source	✓		✓	✓			

Features	GAITRite [67]	Visual3D [7]	NE-Motion [41]	KavaGait [42]	GaitViewer [44]	gaitXplorer [43]	VIGMA (This work)
Format harmonization		△					✓
Feature extraction	△	✓			△		✓
Missing value imputation							✓
Noise filtering		✓					✓
Gait cycle normalization	△	✓					✓
Hierarchical data management		✓		✓		✓	✓
Multivariate feature support		✓		△			✓
Group comparison		△	△	✓			✓
Disease progression	△	△			△		✓
Statistical analysis		✓	△	✓		△	✓
Anomaly analysis		✓	△	✓	△	✓	✓
Access to raw video data	✓						✓
Interactive analysis		△	△	✓	△	✓	✓
Open source							✓

trial information to better understand the data. For instance, videos allow users to identify the reasons behind anomalies or outliers, such as issues with sensors or a misstep on the platform. Similarly, for disease progression analysis, raw trial videos allow users to observe and validate changes in joint angles or ground reaction forces. While many studies have explored visual analytics approaches for analyzing video data [55]–[60], existing literature and tools for gait [7], [42], [44], [61] often overlook the integration of time-series video information. To address this gap, we explore visual analysis and interaction techniques that allow users to correlate raw video data with multivariate and time-series ensemble information.

**Open-access and commercial tools.** Kinovea [62], biomech-Zoo [61], OpenCap [63], and GaitLab [64] are a few open-access tools tailored for gait data. Kinovea supports sports analysis by using videos to capture joint angles and extract a few spatiotemporal parameters (gait speed and stride length), but does not support other data processing tasks or the analysis of more than one trial at a time. OpenCap and GaitLab are tools primarily designed for markerless gait data collection using smartphone camera videos. GaitLab is limited to kinematic gait, whereas OpenCap also estimates force data. However, both tools lack capabilities for data processing tasks such as structured access to patient and trial data, loading external

formats (e.g., C3D, TRC), or further processing (e.g., noise filtering) of collected data. Furthermore, like Kinovea, they also lack the support for analyzing more than one trial at a time. Among these three tools, Kinovea and OpenCap provide access to raw video data. biomechZoo [61], on the other hand, is an open-access toolbox that supports processing, analyzing, and visualizing different types of gait data, including kinetic, kinematic, spatiotemporal, and electromyography. The software consists of MATLAB scripts that perform data processing and two graphical interfaces that support the analysis and visualization of the data. However, MATLAB itself is not an open-source or free-access software. Additionally, biomech-Zoo has complex documentation with numerous functions and generates non-interactive charts, which limits its ability to support tasks related to disease progression and group comparisons.

Commercial gait software [8], [65]–[67] are equipped with proprietary hardware to collect data and are more focused on accurate data collection than data processing, analysis, and visualization tasks. For instance, BioSensics [66] offers real-time gait data collection and monitoring tools like LEGSys and BALANSys for assessing gait steadiness and balance, respectively, but provides little to no data processing or data analysis and visualization support. GAITRite [67], a data collection tool for spatiotemporal parameters, similarly offers minimal

support for data processing, analysis, and visualization. Vicon Nexus [8] is another commercial data collection tool for time-series gait data that supports data processing steps such as feature extraction (e.g., spatiotemporal parameters from ground reaction forces), imputing missing values, and filtering noise, but provides no analysis or visualization support. Templo [65] is a commercial markerless data collection tool that offers single-trial analysis support by visualizing kinematic, kinetic, and electromyography data alongside the corresponding trial video. While Templo supports certain tasks, such as feature extraction and gait cycle normalization, and provides access to raw video data, it is limited to analyzing only one trial at a time.

Among commercial software, Visual3D [7] closely aligns with the proposed framework, offering data processing, analysis, and visualization support for multivariate gait data. However, it supports only a limited range of data formats (C3D and ASCII formats) and does not accommodate TRC or MAT formats. Additionally, it lacks missing value imputation capabilities, provides limited analysis support for time-series gait data with non-interactive charts, and does not support analysis of single-valued gait parameters. The software has lengthy documentation and complex code syntax for processing and analyzing files, as well as a complicated user interface. The proposed system, VIGMA, is an open-access tool that addresses these gaps in existing solutions.

We summarize the commercial and open-access tools, including tools from literature [41]–[44], in terms of data processing, analysis, and visualization features in Table I. In Section VI-C, we present a detailed comparison between VIGMA and a selected subset of these tools that we consider to be the state of the art.

**Computational notebooks.** Computational notebooks, such as Jupyter, have gained widespread popularity for exploratory data analysis and visualization, with users embracing this approach across various levels of coding expertise. However, static visualizations, dependency between cell execution order, and scrolling through too many cells make analysis tedious for practitioners. Recent works [68]–[72] have attempted to address these issues by leveraging code in computational notebooks while visualizing data separately in interactive dashboards, linking them together to combine the strengths of both. Similarly, in the domain of machine learning interpretability, many recent works [73]–[77] have integrated notebooks with interactive visualization dashboards to enhance the interpretation, calibration, and validation of machine learning models.

Working closely with gait experts and surveying diverse gait practitioners across research labs and clinics, we found that they rely on computational notebooks, such as Live Scripts in MATLAB or Jupyter in Python, for data processing tasks. To that end, we adopt a similar approach to these existing studies, handling data processing tasks in computational notebooks and integrating them with the proposed visual analytics system that supports analysis and visualization tasks. The distinction between this work and these studies lies in the fundamental consideration of the problem space. The proposed system is specifically designed for the gait workflows, with unique data processing, analysis, and visualization requirements.

#### IV. SURFACING WORKFLOWS AND REQUIREMENTS

Following design methodologies from prior works [78]–[80], we designed a semi-structured survey in collaboration with two researchers specializing in gait studies (co-authors of this paper) to surface the overall workflow and requirements of practitioners. The survey included a set of questions aimed at collecting insights into how practitioners process gait data, as well as their analysis and visualization needs. The survey was conducted using an online form supplemented by follow-up open-ended questions. Approval for the study was obtained from UIC’s Institutional Review Board (#2016-0933). The primary questions asked of participants were:

- What target population do you collect data from?
- What is the primary goal of your data collection?
- What type of gait data do you typically work with?
- What are the formats of data you collect?
- What data processing tasks do you perform?
- What data analysis tasks do you perform?

In addition, we also asked participants to evaluate current visualization designs used for interpreting gait data. We had a total of 23 participants for the survey, with diverse job titles and from different research labs and clinics. The majority of survey responses, comprising 15 out of 22 (65.2%), were from researchers. Among these 15 respondents, six also identified as clinicians and six as students. Completing the survey took participants approximately 25 to 30 minutes. The survey questionnaire, participant responses, and the corresponding requirements extracted from these responses are summarized in Table I of the appendix. The eleven visualizations presented to the participants and the evaluations from the participants are presented in Fig. 1 and Table II of the appendix, respectively.

After the survey was completed, we inspected the results to surface a set of requirements for the framework. This process was conducted through meetings between the authors of this paper, which included both visualization researchers and domain experts. We derived three data processing requirements (**R1** - **R3**) and six data analysis & visualization requirements (**R4** - **R9**). To address the needs of a broad user group comprising gait practitioners from different research labs and clinics, we sought to identify their visual encoding preferences. Building on existing studies [81]–[83] that emphasize comprehension, usability, and user preferences in visualization design, we presented the participants with eleven visualizations (Fig. 1 in the appendix). These visualizations were selected after a review of existing literature and gait analysis and visualization tools [32], [42], [44], [84]–[87]. We asked participants to evaluate them in terms of usability and understandability. We believe such practices are valuable in design studies of visualization, as supporting a diverse group of domain experts with limited visualization experience requires using visual designs that prioritize usability and comprehension. Furthermore, existing works [88]–[90] highlight that relying on familiar designs rather than novel glyphs is more successful with wider audiences, especially with healthcare experts. Thus, we selected familiar visual designs among experts through a literature review and then refined the selection for the proposed system based on the survey results.

The findings on visual design preferences are outlined in Section IV-A, followed by the data processing, and data analysis and visualization requirements extracted from the survey in Section IV-B.

#### A. Findings on Visual Design Preferences

The eleven visualizations presented to the survey participants support displaying time-series gait data (e.g., joint angles, ground reaction forces) as well as single-valued spatiotemporal parameters (e.g., gait speed, swing time). We asked the participants to evaluate these visualizations in terms of usability and understandability. Based on the mean user scores, along with the additional comments provided by the participants, we derived the following findings:

- To analyze time-series gait data, existing literature and tools almost solely rely on line plots (Fig. 1-A-F in the appendix). The survey participants were also found to be accustomed to using line plots, and, on average, scored them comparably higher than other types of visualizations.
- To analyze the distribution of single-valued parameters such as sociodemographic or spatiotemporal gait parameters, the participants expressed preference for box plots (Fig. 1-I in the appendix) over strip plots (Fig. 1-G, H in the appendix), finding strip plots to be more complex and less intuitive.
- For displaying values and making comparisons involving spatiotemporal parameters, participants found the radar plot (Fig. 1-J in the appendix) to be highly useful. This visualization method is particularly beneficial for representing single-valued gait parameters relevant to rehabilitation or disease progression across multiple trials or when comparing different patient groups, such as stroke patients and older adults.

We leveraged the visualization preferences of the survey to tailor familiar solutions (i.e., line, box, and radar plots) that satisfy the requirements of the framework. The design of visual encodings and interactions is discussed in detail in Section V-B.

#### B. Requirements

Here, we describe the extracted data processing requirements in **R1 – R3**, followed by the data visualization and analysis requirements in **R4 – R9**.

**R1. Data standardization.** Gait data originates from different hardware and software systems like Vicon Nexus [8] and TEMPLO [65], generating diverse formats such as MAT, C3D, TRC, TXT, and CSV, which introduces challenges for collaboration and data sharing. To address these, the framework should support different data formats and harmonization features. Additionally, it should support feature extraction tasks, such as deriving joint angles, step times, spatiotemporal parameters, and fall risk predictions from combinations of motion, ground reaction forces, and joint data. Lastly, the framework should offer data preparation support for imputing missing values, filtering noise, and normalizing data to align gait cycles and ensure consistency across trials.

**R2. Data management.** Data is gathered from a diverse range of patient groups, such as stroke patients, older adults, and cerebral palsy patients, across various research labs and clinics. To conduct efficient gait data analysis, practitioners need to have access to two critical pieces of information: (1) patient groups and (2) individual patient trials. This enables effective comparisons between different patient groups or various trials of the same patient. Consequently, the proposed framework should support accessing patient data in this hierarchical manner. Furthermore, for analysis tasks, users often need to refer back to the raw trial videos. Therefore, the framework should support multimodal data access, ensuring on-demand availability of raw trial videos.

**R3. Computational notebooks.** Practitioners heavily rely on computational notebooks, such as Live Scripts in MATLAB or Jupyter in Python, for their data processing needs. Notebooks allow them to perform data processing tasks, revisit previous steps, check processed data, identify errors, and re-execute steps to ensure the correctness of the processed data. Additionally, notebooks are easily shareable, fostering collaboration. However, practitioners' programming expertise is often limited. Thus, the framework should support performing data processing tasks (**R1**, **R2**) through short one-liner functions within computational notebooks.

**R4. Multivariate characteristics.** Gait data is collected in diverse forms, including ground reaction forces, 2D/3D motion, joint angles, electromyography, joint moment, spatiotemporal parameters, and more. These data can be categorized into two primary types: time-series and single-valued parameters. To facilitate effective gait analysis, the framework should possess the capability to analyze and visualize both types of gait characteristics of patients.

**R5. Group comparison.** Facilitating the comparison of different groups of gait trials allows for a detailed analysis of key characteristic differences between patient groups. For instance, comparing gait data of stroke patients with that of healthy older adults can reveal significant differences, allowing practitioners to focus on improving specific gait attributes in stroke patients. These types of analyses are crucial in gait studies. Therefore, the framework should support the aggregation of trials and the comparison between ensembles.

**R6. Disease progression.** Tracking rehabilitation or disease progression of patients over time is a key aspect of gait studies. For instance, practitioners often monitor key gait characteristics of stroke patients over a period of 6-12 months, adjusting rehabilitation strategies based on improvements or deteriorations in the patient's condition. This approach is similarly applied to injured athletes, cerebral palsy patients, and others. Consequently, the framework should enable seamless analysis of rehabilitation and disease progression of patients.

**R7. Statistical measures.** The analysis of statistical measures of gait data is important in acquiring insights into data distribution and facilitating comparisons between different patient groups. In addition, it enables the assessment of how a patient trial deviates from the distribution of a different group or within its respective group. To support these aspects, the framework should enable the analysis of statistical measures, including confidence interval for time-series gait data and

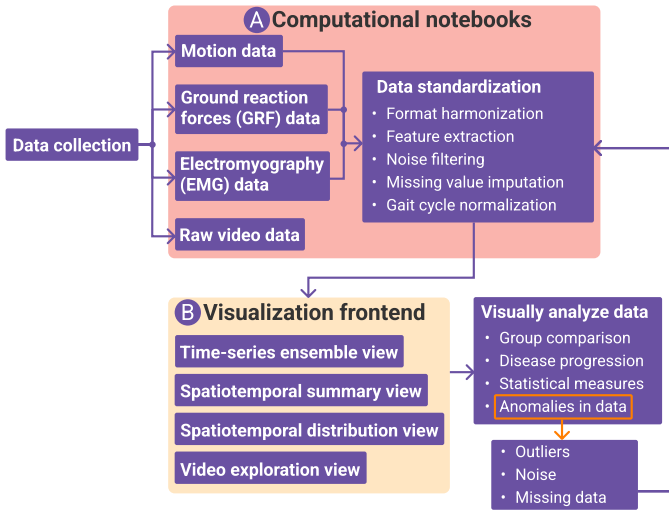


Fig. 3. Overview of the workflow. The framework comprises two main components: **A** computational notebooks and **B** visualization frontend. The notebooks facilitate processing gait data (e.g., motion data) by performing standardization steps (e.g., feature extraction, noise filtering) before sending it to the visualization frontend. Users can interact with the frontend to select ensembles of trials and generate charts for time-series and spatiotemporal gait data using time-series ensemble view, spatiotemporal summary view, and spatiotemporal distribution view and have access to raw video data through video exploration view. If anomalies are identified in the data, users can return to the notebooks to correct the data and then visualize the corrected data in the frontend.

mean, median, percentiles, and standard deviation for single-valued parameters.

**R8. Anomaly analysis.** Identifying anomalies in gait data is crucial for understanding unusual patient behaviors and ensuring data accuracy. For instance, detecting outliers in the data can help tailor training or clinical interventions to address specific patient needs. Additionally, identifying erroneous data (e.g., missing data, noise) prompts necessary corrective actions to maintain data integrity. The framework should support the detection of outliers and erroneous data, as well as provide corrective measures to fix the errors.

**R9. Interactive analysis.** The framework should support selection, filtering, highlighting, overview-first and details-on-demand, as well as cross-view interactions, to enable human-in-the-loop workflows [91] for analysis and visualization tasks (**R4** – **R8**). For multivariate analysis (**R4**), users should be able to **select** different gait variables and parameter configurations. For group comparisons (**R5**) or tracking disease progression (**R6**), users should be able to **filter** from the entire dataset, **selecting** ensembles of trials. To identify and analyze anomalies (**R8**), users should have the ability to **highlight** specific trials or single-valued parameters in the visualizations and view **details** such as patient IDs and trial numbers on demand. Additionally, the framework should provide access to **detailed** numeric values, such as mean, median, and confidence intervals, alongside the visualizations (**R7**). Lastly, **cross-view interactions** between visualizations would enable users to explore relationships among different gait variables, facilitating more in-depth insights into tasks such as group comparisons (**R5**), disease progression (**R6**), and anomaly analysis (**R8**).

```
import vigma
# Format harmonization: MAT to CSV
df = vigma.matToCSV
# Feature extraction: motion to joint angle
df = vigma.motionToJointAngle
# Data preparation: missing value imputation
df = vigma.mice_impute
# Data preparation: filtering noise
df = vigma.filter_data
# Data management: save data
vigma.save
```

Listing 1. VIGMA Python API to perform data processing tasks.

## V. THE VIGMA FRAMEWORK

VIGMA was designed based on the requirements collected in Section IV. We followed a human-centered design approach, iteratively prototyping and implementing the system while gathering continuous feedback through bi-weekly one-hour in-person meetings with two gait researchers (co-authors of the paper). Their feedback primarily helped us make the Python API less code-intensive and more comprehensible to domain experts, improve the overall system’s layout, and select encoding designs and user-friendly interactions for the visualization frontend.

VIGMA has two main components: (1) A Python library that can be accessed in computational notebooks (Fig. 3-A), and (2) a visualization frontend (Fig. 3-B). Data processing tasks are carried out using computational notebooks. The processed data are then accessed and visualized using the frontend. Fig. 3 details the interactions between the two components. Sections V-A and V-B detail the data processing and visual design of the framework, respectively, followed by implementation details in Section V-C.

### A. Data Processing

To streamline gait data processing, our framework contains a Python library that simplifies data tasks into single-line function calls that can be used in computational notebooks (**R3**). Example of the VIGMA Python API is illustrated in Listing 1.

The API facilitates data standardization (**R1**) by converting formats such as TRC, MAT, and C3D into structured dataframes, which can then be saved as CSV files. It supports feature extraction of joint angles, step times, spatiotemporal parameters, joint moments, and fall risk predictions derived from motion data, ground reaction forces, and center of mass. Additionally, it offers support for missing value imputation, noise filtering, and normalization of trials by gait cycles. The API allows (1) saving trial data using a hierarchical file structure that stores it by patient groups and individual patients and (2) loading it into the visualization frontend (**R2**). Users can access raw trial videos (**R2**) alongside the processed data through the visualization frontend.

### B. Visual Design

The frontend of VIGMA is composed of three different views: (1) time-series ensemble view, (2) spatiotempo-



ral summary view, and (3) spatiotemporal distribution view. The spatiotemporal summary and distribution views can be alternatively swapped altogether with the video exploration view using a checkbox from the control panel (Fig. 1-E). The video exploration view provides access to raw trial videos that are linked to the trials from the time-series ensemble views (R2). The frontend also includes a control panel that allows seamless access and filtering of patients' trial data through the hierarchical structure, organized by patient and patient group (Fig. 4) (R2, R9). In addition, the control panel lets users visualize single or dual groups of selected ensembles and choose chart configurations, including parameter(s) to visualize, limb side, and gait cycle (R9) (Fig. 1-E). All visualizations in the different views are linked via cross-view interactions that allow highlighting of trials for different gait variables for more in-depth analysis (R9). We provide design justifications for the three views and cross-view interactions. In addition, we report all the alternative design choices in Fig. 2 in the appendix section. The three views of the visualization frontend enable users to compare between two groups of ensembles (R5), analyze disease progression (R6), explore statistical measures (R7), and find and highlight anomalies (R8).

1) *Time-series ensemble view*: The frontend contains four containers for time-series ensemble views (Fig. 1 F1-F4), which display line charts for time-series ensemble data based on different parameter configurations selected in the control panel. Users can select any type of time-series data (e.g., motion data, joint angles, ground reaction forces) to display in the time-series ensemble views (R4). The view allows the creation of two different ensemble groups and a comparison of their time-series characteristics (R5). Users can also select trials of the same patient from two different time periods to assess the progression of their time-series characteristics (R6). The view displays the ensemble mean of selected trials along with all individual trials (Fig. 5-A1) or the confidence interval (Fig. 5-A2) of the ensemble (R7). Hovering over a line highlights it and reveals the corresponding patient and trial id (Fig. 1-I1) with a tooltip, facilitating the identification of outliers and anomalies (R8, R9). In case of line with confidence intervals, hovering displays a tooltip showing the numeric values of the mean, lower bound, and upper bound at a specific time point for the selected gait parameter (Fig. 1-I2)

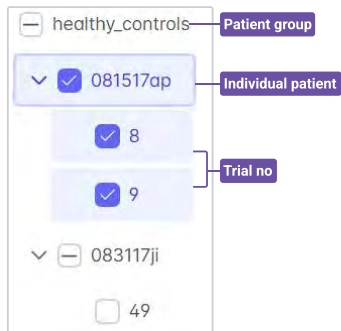


Fig. 4. Accessing data through the control panel of the visualization frontend with each trial hierarchically organized by patient ID and patient group.

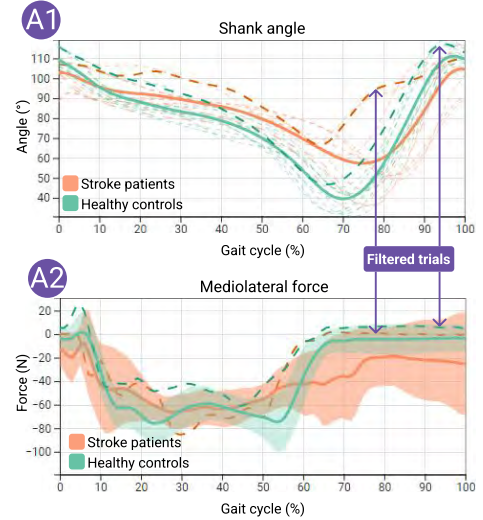


Fig. 5. Time-series ensemble view. Displays the ensemble mean (green and orange solid lines) along with - A1 individual trials' data (faded dashed lines) or A2 the confidence interval - for two different groups (i.e., stroke patients and healthy controls). Users can filter and highlight a trial in A1 by clicking on a faded dashed line, which will then appear as a bold dashed line within its view, other time-series ensemble views (e.g., A2), and the spatiotemporal summary view (Fig. 6). The filtered trials are also highlighted in the spatiotemporal distribution view (Fig. 7) with rectangular brushes, indicating their range of values for each single-valued spatiotemporal parameter. Conversely, adjusting the rectangular brushing on any box plot in the spatiotemporal distribution view will highlight a different set of trials in the time series ensemble views (A1, A2) and spatiotemporal summary view (Fig. 6).

(R7, R9).

*Justification.* Initially, we considered displaying multiple variables within a single chart (Fig. 2-A1 in the appendix) and using side-by-side plots for group comparisons (Fig. 2-A2 in the appendix). However, experts found it difficult to distinguish multiple variables (e.g., vertical force, mediolateral force) within one chart, and comparison using side-by-side plots was challenging for them. As a result, we transitioned to four separate time-series ensemble views (Fig. 1 F1-F4), each dedicated to a single variable. Each view allows comparisons between up to two groups of trials (e.g., stroke, healthy). Additionally, our initial approach displayed only the ensemble mean of a variable (Fig. 2-B1 in the appendix). However, experts preferred to see all individual trials to better identify anomalies, leading us to incorporate both individual trials and the ensemble mean (Fig. 5-A1). To further facilitate finding anomalies, we added a hover functionality that highlights a trial and displays its corresponding patient and trial ID (Fig. 1-I1). For tasks not focused on finding anomalies but rather on understanding distribution patterns, experts requested a less cluttered view. In response, we introduced a mode displaying the mean with a confidence interval (Fig. 5-A2) and a tooltip that provides numerical values for the mean and confidence bounds (Fig. 1-I2).

2) *Spatiotemporal summary view*: The ensemble mean of the spatiotemporal gait parameters is displayed using a radar chart in this view (Fig. 6). This view allows users to effectively analyze multivariate spatiotemporal gait parameters (R4) and compare them between two groups (R5). Additionally, users can select data from the same patient across two time periods



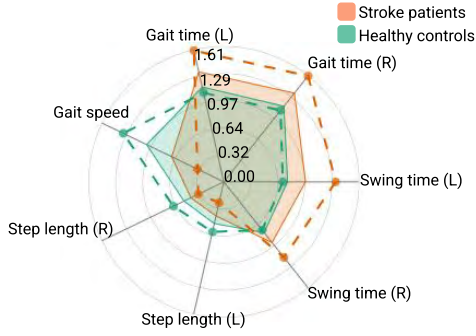


Fig. 6. Spatiotemporal summary view. Displays the ensemble mean (opaque green and orange areas) of spatiotemporal parameters for selected lists of trial data for two different groups. When users filter trials by clicking on faded dashed lines from the time series ensemble view (Fig. 5) or brushing the range of a spatiotemporal parameter from one of the box plots in the spatiotemporal distribution view (Fig. 7), the filtered trials get highlighted with bold dashed lines in this view.

to examine the progression in spatiotemporal parameters over time (R6). Hovering over a parameter in the radar chart displays a tooltip with the numeric mean of the two ensembles for that parameter (Fig. 1-I3) (R9).

*Justification.* The experts supported the survey results favoring a radar chart, noting that it provides a quick overview of how two ensembles differ across spatiotemporal parameters. While box plots (Fig. 7) show distribution differences, the radar chart (Fig. 6) enables easier comparison of multiple variables at one glance. Experts also found it more intuitive than box plots for assessing disease progression for spatiotemporal parameters. Initially, we considered displaying more than two ensembles in the radar chart (Fig. 2-C1 in the appendix), but the experts noted that too many colors made the chart difficult to interpret and mentioned that comparing two ensembles was sufficient for their purposes. We also added the hover functionality (Fig. 1-I3), as the experts wanted to see exact numeric values on demand.

3) *Spatiotemporal distribution view:* The distribution of multivariate spatiotemporal parameters (R4), including min, max, median, upper, and lower percentiles (R7), for the selected ensembles of trials is displayed using dual box plots in this view (Fig. 7). Users can also compare the distribution of two patient groups using this view (R5). Hovering over any box displays a tooltip with the statistical distribution of

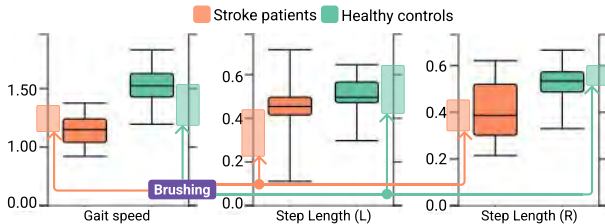


Fig. 7. Spatiotemporal distribution view. This view features dual box plots for each spatiotemporal parameter, displaying the distribution for an ensemble of selected trials across two groups. Users can filter each plot for either group using rectangular brushes. Filtering one box plot will subsequently update the other box plots, as well as the time-series ensemble view (Fig. 5) and the spatiotemporal summary view (Fig. 6).

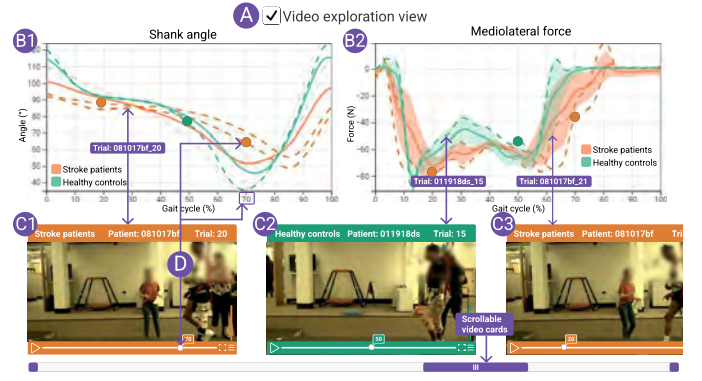


Fig. 8. Access to raw trial videos C1 - C3 is provided through a scrollable video exploration view. Users can toggle between spatiotemporal views (Fig. 6, Fig. 7) and the video exploration view using a checkbox A. Videos appear in the video exploration view when users click on and highlight trials in the time-series ensemble views B1, B2. The video cards C1 - C3 are color-coded by patient group and display patient and trial id, and patient group information in the header. Videos are bi-directionally linked with the time-series ensemble views, with a circular marker D appearing in the time-series ensemble views to indicate the corresponding time in the video.

the selected spatiotemporal parameter (Fig. 1-I4) (R7, R9).

*Justification.* The experts emphasized that box plots are essential for detailed statistical analysis of spatiotemporal parameters and are the most familiar visualization among gait practitioners. Initially, we displayed a single box plot per parameter (Fig. 2-E1 in the appendix), but the experts requested the ability to compare distributions between two groups, leading to the adoption of dual box plots (Fig. 7). To meet their need for numeric values on demand, we also added a tooltip (Fig. 1-I4) that appears on hover.

4) *Cross-view interactions:* The frontend of VIGMA features several cross-view interactions, enabling interactive analysis between multiple coordinated views [92].

i) *Clicking.* Users can highlight one or more trials by clicking on them in the time-series ensemble view, which renders them as bold dashed lines (Fig. 5-A1, A2). The highlighted trials will also appear in bold dashed lines in all other time-series ensemble views and the spatiotemporal summary view (Fig. 6). Additionally, the spatiotemporal distribution view will display a brush indicating the range of values for the highlighted trials (Fig. 7). Clicking on a trial in the time-series ensemble view also provides access to the raw video for that trial within the selected gait cycle timeframe (Fig. 8) (R2).

ii) *Brushing.* The highlighting of trials across views works bi-directionally. If users brush a range in any box plot within the spatiotemporal distribution view (Fig. 7), the corresponding trials within the brush range will be highlighted in all time-series ensemble views (Fig. 5). Additionally, the highlighted ensemble mean in the spatiotemporal summary view will update accordingly, displayed as bold dashed lines (Fig. 6).

iii) *Animated transition.* Users can play or pause raw videos from the video exploration view, which triggers a circular marker to appear on the time-series ensemble views for the corresponding trial (Fig. 8). As the video plays, a circular marker dynamically moves across all time-series ensemble views corresponding to that trial.

*Justification.* The experts wanted to analyze correlations between multivariate gait parameters across different views to gain deeper insights into tasks such as disease progression, outlier detection, and group comparisons. To support this, we implemented a click interaction within the time-series ensemble view. While hovering provided only temporary focus, clicking allowed users to highlight selected trials persistently across all views (Fig. 5-A1, A2). For analyzing the distribution of spatiotemporal parameters and exploring their behavior across other views, we incorporated a brushing interaction in the dual box plots (Fig. 7). To provide access to raw videos of trials, we initially considered using pop-ups (Fig. 2-F1 in the appendix) linked to clicked lines in the time-series ensemble views. However, the experts found pop-ups distracting because they required switching focus to a separate window, breaking the connection between the video and the corresponding visualizations. Since the videos are tied only to the time-series parameters, instead of popups, we designed a video exploration view (Fig. 8) that displays videos in a scrollable format in lieu of the spatiotemporal summary and distribution views. Lastly, the experts wanted to maintain context between the video and the corresponding time-series data. We introduced circular markers (Fig. 8-D) that move along the time-series ensemble views, synchronized with the playback of the trial video, allowing users to visualize dynamic changes in time-series parameters as the video progresses.

### C. Implementation

The system was implemented using Flask for the backend and React with D3.js [93] for the frontend. The computational notebooks employ a custom Python library we developed, which utilizes numpy, Pandas, SciPy, C3D, Scikit-learn, and plotly.

## VI. EVALUATION

We evaluated the VIGMA framework through three usage scenarios created in collaboration with domain experts and experts' quantitative and qualitative feedback on the framework's understandability and usefulness.

### A. Usage Scenarios

We designed the usage scenarios in collaboration with two gait researchers (co-authors of the paper), using the necessary data from their research lab. The lab focuses on fall prevention, motor control, and stroke rehabilitation, collecting multivariate gait data from a diverse range of patients, including children, young adults, healthy older adults, and stroke patients. Through in-person meetings with the researchers, we identified common practices performed by gait practitioners across various labs and clinics, including:

- Addressing data quality issues due to hardware limitations, necessitating the identification and correction of such errors.
- Analyzing data from patients with disorders (e.g., stroke, parkinson) at intervals (e.g., 6 or 12 months) to identify rehabilitation progress.

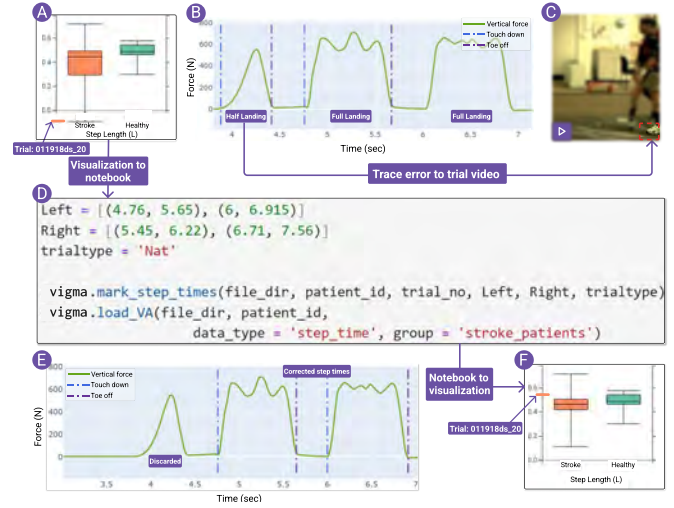


Fig. 9. Usage scenario 1. **A** User identifies an error of negative step length in one of the stroke patient's trials from the spatiotemporal distribution view. **B** By examining vertical force and step times (touchdowns and toe-offs), they determine the cause of the error to be half landing on the platform and **C** validate it against the trial video using video exploration view. Afterwards, **D** user corrects the error by using the VIGMA library in a notebook, **E** checks if the half landing is properly discarded and **F** returns to the visualization to see the updated value.

- Comparing data from patients with disorders to data from healthy subjects to identify deviations.

Considering these common practices, we collected data from twenty subjects, among which ten were healthy subjects (aged from 61 to 76 years, mean  $65.50 \pm 3.91$ ; body mass  $84.72 \pm 14.49$  kg; height  $1.68 \pm 0.08$  m; 6 males; 4 females) and the other ten were stroke patients (aged from 52 to 69 years, mean  $61.20 \pm 4.73$ ; body mass  $85.24 \pm 18.40$  kg; height  $1.72 \pm 0.09$  m; 7 males; 3 females). The data of stroke patients was collected at the time of diagnosis and again after a 6-month interval. For the scenarios, we considered the following types of data: 3D motion, joint angles, ground reaction forces, and spatiotemporal parameters. 3D motion data and ground reaction forces were collected using motion capture cameras and force plates during regular walking trials on a 7-meter walkway, as illustrated in Fig. 2. The collected data, originally in TRC, MAT, and C3D formats, were standardized and converted to CSV using our framework. In addition, we utilized computational notebooks to extract joint angles from motion data, step times from ground reaction forces, and spatiotemporal parameters from joint angles and step times.

The collected data was then used in the following usage scenarios: (1) identifying and correcting erroneous data, (2) stroke rehabilitation analysis, and (3) comparing healthy subjects and stroke patients. Stroke patients and healthy subjects, whose data were collected, were informed about the purposes of the study and provided written informed consent. The study was approved by UIC's IRB. Data collected for healthy controls and stroke patients fall under IRB protocol #2016-0887 and #2016-0933, respectively.

*1) Usage scenario 1: Identifying and correcting erroneous data:* Anomalies, such as outliers and errors, are common occurrences in the process of collecting gait data. For the

first usage scenario, we began by selecting ten stroke patients and ten healthy patients as two separate groups using the control panel and visualizing the spatiotemporal parameters using the spatiotemporal distribution view. We observed an instance of negative step length (left foot) for one of the trials from the stroke patients and identified the erroneous patient and trial id by brushing the box plot (Fig. 9-A) (R8). Analyzing the vertical force for that trial, we identified a half landing at the beginning of the trial (Fig. 9-B) and validated our speculation by tracing it to the original trial video using the video exploration view (Fig. 9-C). To correct the issue, we used our Python API in a notebook to mark step times by discarding the half landing and loaded the corrected data to the visual analytics system (Fig. 9-D, E) (R1). Returning to the spatiotemporal distribution view, we then identified the correct step length for the left foot for that trial (R8).

This usage scenario demonstrates how users can leverage notebooks and the visualization interface to seamlessly navigate between data processing and visualization tasks.

2) *Usage scenario 2: Stroke rehabilitation analysis:* The second usage scenario demonstrates the system's capability to analyze stroke rehabilitation progress over time, a key focus in motion and gait studies (R6).

First, we selected the data using the control panel for one stroke patient at two different time points: the first trial and after six months. The patient was instructed to reduce braking force and gait speed to minimize fall risk during walking. We generated three time-series ensemble views for anterior-posterior force (Fig. 10-A), shank angle (Fig. 10-C), and trunk angle (Fig. 10-D), as well as a spatiotemporal-summary view (Fig. 10-B) to compare the spatiotemporal parameters over the six-month period. Analyzing the views, we observed a reduction in the patient's braking force, particularly during the initial phase of the gait cycle (Fig. 10-A). Additionally, the overall gait time and swing time for both feet increased after six months, resulting in a reduced gait speed. This indicated an overall improvement in the patient's gait, reducing the risk of falls. Furthermore, analyzing the shank and trunk angles, we observed an increase in both knee flexion (Fig. 10-C) and trunk flexion (Fig. 10-D), indicating improved flexibility and stability in these limbs.

3) *Usage scenario 3: Comparing healthy subjects and stroke patients:* Practitioners frequently compare gait characteristics across different groups of patients to identify key differences and to analyze how an individual patient's data aligns or differs with that of a specific patient group. Our third usage scenario specifically examines this aspect in detail (R5).

We began by comparing the left and right foot angles in a right-sided paretic stroke patient (Fig. 11-B) to a healthy control (Fig. 11-A). We observed that the healthy control shows consistent patterns for both the left and right foot, suggesting a balanced and symmetrical gait. However, the stroke patient exhibited clear asymmetry between their paretic and non-paretic sides, and we were able to analyze the degree of asymmetry in the foot angles.

For the next part, we merged data from five right-sided paretic stroke patients and compared their right anterior-

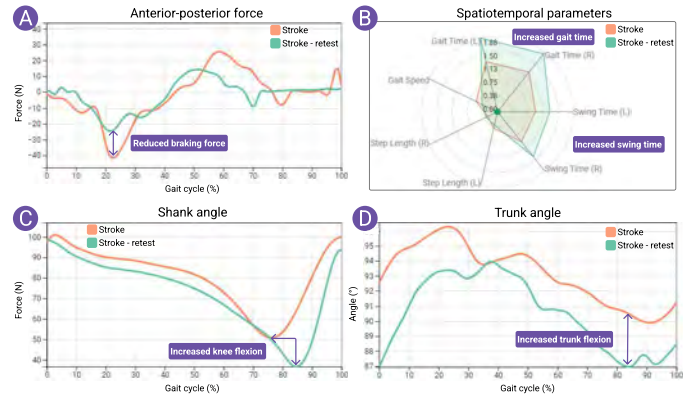


Fig. 10. Usage scenario 2. Identifying the rehabilitation progress of a stroke patient over time. Orange represents data at the time of the stroke, while green represents the same patient's data after 6 months. The patient received instructions to increase gait time and reduce braking force to minimize fall risk. Improvements in the post-6-month data show A) reduced braking force, B) increased gait and swing time, and reduced gait speed. Additionally, the patient exhibited more C) knee and D) trunk flexion, indicating improved balance and stability in their gait.

posterior force (Fig. 11-C) and spatiotemporal parameter distribution (Fig. 11-D) with those of five healthy controls. We observed that stroke patients exhibited a flatter anterior-posterior force curve with less pronounced peak (propulsive force) and trough (braking force) than healthy controls. The lower propulsive force in stroke patients suggests that they have difficulty in generating forward momentum during walking, which impacts their gait speed. The spatiotemporal distribution view supports this hypothesis, revealing that the gait speed distribution for stroke patients is significantly lower than that of healthy controls.

Lastly, we analyzed the outliers in stroke trials that exhibited higher gait speeds by examining the box plot of gait speed in Fig. 11-D (R8). We discovered that these trials had higher step lengths compared to most stroke patients by visualizing the distributions of left and right step lengths (Fig. 11-D). Additionally, we observed that their right anterior-posterior force (Fig. 11-E) and foot angles (Fig. 11-F) were more similar to those of healthy controls than to other stroke patients. By tracing these trials to their corresponding videos, we identified that they all belonged to the same patient. Further analysis of these trial videos revealed that this patient had already improved their gait characteristics to a certain extent and required less intensive rehabilitation strategies.

## B. Experts' Feedback

After creating the usage scenarios, we engaged with five experts from different research labs and clinics to get their feedback on VIGMA. The five experts are not co-authors of the paper. Table II summarizes their background.

Each expert was exposed to all three usage scenarios, with each session lasting approximately 1.5 hours to complete. To save time, three separate instances of VIGMA, preloaded with the required data for each usage scenario, were opened before starting the session. The experts were asked to run the usage scenarios themselves by executing the computational



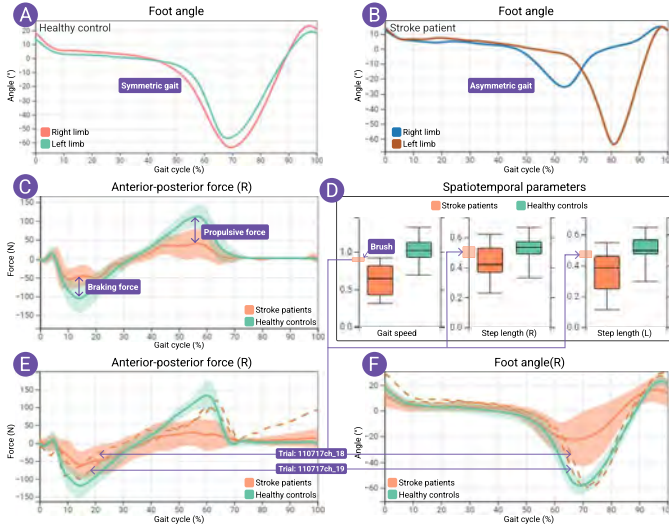


Fig. 11. Usage scenario 3. Comparing the foot angle of **A** a healthy older adult with **B** a right-sided paretic stroke patient to identify the magnitude of asymmetry in the stroke patient's gait. The user then ensembles trials from five right-sided paretic stroke patients and five healthy controls to compare their **C** right-sided anterior-posterior force and **D** spatiotemporal parameters. Additionally, the user brushes the box plot for stroke patients with higher gait speeds and highlights the corresponding trials for **E** right-sided anterior-posterior force and **F** right-sided foot angle. The user finds that the highlighted trials belong to the same stroke patient whose gait is more similar to those of healthy controls, suggesting they require a less intensive rehabilitation strategy.

notebooks and interacting with the visualization frontend, receiving assistance if they encountered any issues. Two of them participated in the study in person, while the other three joined via Zoom meetings with desktop sharing. It is important to note that the experts who tested the usage scenarios and provided feedback were different from the two researchers who collaborated in the creation of the scenarios.

We gathered both qualitative and quantitative feedback from each expert. The sessions for testing the usage scenarios were open-ended, allowing the experts to provide qualitative feedback both during and after the sessions. At the end of the session, the experts were asked to complete an online questionnaire to provide quantitative feedback.

1) *Qualitative feedback*: Overall, our framework received positive qualitative feedback from the experts. One of the senior clinicians and faculty members affirmed: “As a clinician I feel the visualization tool is easy to understand and very useful to plot and visualize data. I would definitely use this to plot my information.” The clinician found the studies to be enlightening, particularly in demonstrating how the different views could help analyze the rehabilitation progress of stroke patients and design more effective rehabilitation strategies. Another senior clinician echoed this sentiment, stating, “the tool comes with awesome flexibility. Very helpful in detailed analysis.” This clinician was particularly impressed by the flexibility of processing and analyzing various types of multivariate gait data using the system.

Ongoing PhD students and researchers found the tool to be equally useful. One of them remarked “the tool can serve in both clinical and research capacities.” Another student noted

TABLE II  
PARTICIPANTS INVOLVED IN THE EVALUATION. EACH EXPERT TESTED ALL THREE USAGE SCENARIOS AND LATER PROVIDED FEEDBACK. THE EXPERTS ARE NOT CO-AUTHORS OF THE PAPER. BACKGROUND ABBREVIATIONS: F - FACULTY, C - CLINICIAN, R - RESEARCHER, S - STUDENT.

Background	Years in field	Knowledge in gait (1-5)	Education level
F, C, R	5	4	PhD
S	2	3	PhD (Pursuing)
F, C	10	4	PhD
C, R	1	2	Masters
C, S	6	4	PhD (Pursuing)

“a system like this will be extremely useful for both clinicians and researchers. From a clinician perspective, it will expand our ability to perform detailed gait analyses without requiring an expert biomechanist or computer engineer - expanding also the number of people who can be positively impacted by this system. Additionally, as a PhD Candidate, I think that it will help people like me work smarter and expand our ability to analyze more data and obtain results which are more impactful.” We received highly positive remarks for the inclusion of videos in the system. One researcher commented that “having access to videos in the system is very helpful in detecting and validating gait abnormalities or deviations.”

We received mixed feedback regarding the computational notebooks used for data processing tasks, with one of the experts mentioning, “although I have very little experience in coding, the notebook side of things seemed very straightforward and easy to me.” On the other hand, one of the clinicians stated “down the road, I think it would be useful to have buttons to represent the notebook functions for individuals with less computer science/engineering background.” Another clinician emphasized the need for proper instructions to effectively use the API, stating, “this would require training or step-by-step instructions to ensure understanding.”

2) *Quantitative feedback*: In addition to qualitative feedback, we asked the experts to complete an online questionnaire rating the usefulness and understandability of the whole framework and individual components on a Likert scale. The results are shown in Fig. 12.

Feedback was positive, especially regarding usefulness, with all ratings falling between 4 and 5. Understandability ratings varied more, with computational notebooks receiving lower scores compared to other components. This aligns with qualitative feedback from some experts, who noted a desire for the computational notebooks to be more simplified or to include proper documentation with examples for easier use of the API. Based on feedback from the experts, we have made the complete API documentation available on our GitHub page, including examples of one-line function calls to facilitate easy replication of data processing tasks.

### C. Comparison With the State of the Art

We also compared VIGMA with existing tools across data processing, analysis, and visualization features. (see Table I). Here, we discuss tools selected from the table based on whether they partially or fully support at least 35% of the

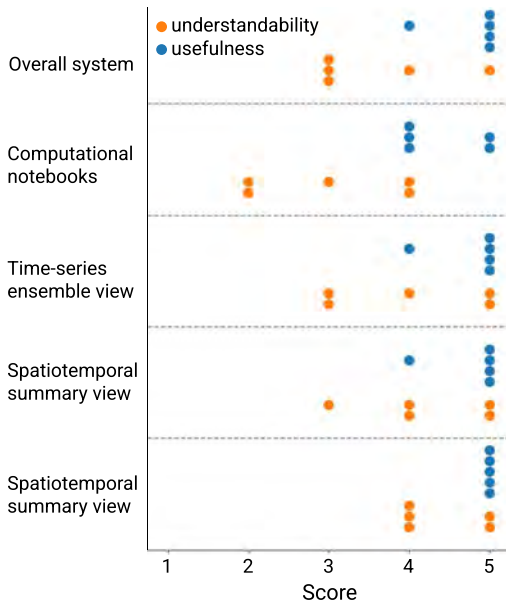


Fig. 12. VIGMA’s usefulness and understandability scores from experts on a Likert scale from 1 to 5.

features. The tools include: (1) bioMechZoo [61], a MATLAB toolbox for processing, analyzing, and visualizing multivariate gait data; (2) Vicon Nexus [8], a multivariate gait data collection tool that utilizes markers, motion cameras, and force plates, providing some data processing and analysis capabilities; (3) TEMPLO [65], a markerless gait data collection tool allowing for single-trial analysis of different gait data types; (4) Visual3D [7], an all-in-one data analysis tool for processing, analyzing, and visualizing various types of multivariate gait data; (5) KAVAGait [42], a gait analysis tool for storing and analyzing implicit knowledge of gait data such as ground reaction forces and spatiotemporal parameters through visualizations; and (6) Kinovea [62], a video analytics tool for gait analysis, primarily designed for sports.

Regarding data processing requirements, only Visual3D partially or fully meets most of the criteria. It offers limited format harmonization (C3D, ASCII), and while it supports feature extraction, noise filtering, and gait cycle normalization, it does not support missing value imputation (R1). Additionally, the data processing steps are performed through a complex interface and do not support the integration of computational notebooks based on Python or MATLAB (R3). Among the other tools, only bioMechZoo offers comprehensive support for feature extraction, missing value imputation, noise filtering, and gait cycle normalization (R1), and its MATLAB-based code can be integrated with notebooks (R3). However, it only supports MAT files for data formats (R1) and lacks organized access to patient data through a visual interface (R2). Additionally, both of these systems have lengthy documentation and complex code syntax for processing data. Finally, among the six selected tools, only TEMPLO and Kinovea provide access to raw video data (R2).

Concerning data analysis and visualization requirements, Visual3D supports statistical (R7) and anomaly analysis (R8) through non-interactive charts. However, limited interactions

(R9) makes it inadequate for performing patient group comparisons (R5) or disease progression (R6). bioMechZoo faces the same limitations. KAVAGait offers interactive analysis (R9), but is limited to analyzing only a specific subset of gait data types (R4). KAVAGait also supports patient group comparisons (R5), but the analysis is limited to spatiotemporal parameters (R4). Lastly, of the six tools considered, only Kinovea is open source, underscoring a significant gap in the availability of open-source tools to drive research and clinical gait analysis.

## VII. CONCLUSION

This paper introduces VIGMA, an open-access visual analytics framework designed for multivariate gait analysis that integrates computational notebooks to facilitate data processing tasks and a frontend for visualization and data analysis tasks. The Python API offers one-liner function calls to simplify complex data processing tasks, and then seamlessly integrates processed data into the visual frontend, enabling comprehensive analysis and visualization of multivariate gait data. The usage scenarios demonstrate how the system achieves its goal of streamlining the integration between data processing, and data analysis and visualization, while efficiently handling common gait analysis tasks such as tracking disease progression and performing group comparisons. VIGMA’s open-access framework aims to foster greater collaboration between research labs and clinics by offering a one-stop solution for gait workflow.

**Limitations.** While promising, the current state of VIGMA presents some limitations. Some clinicians who participated in the usage scenarios highlighted that, although they found the one-liner function calls useful for data processing tasks, they would prefer something even simpler, and added the need for detailed documentation. To address this, we have made the documentation available on GitHub, including examples for easy replication. Additionally, the current usage scenarios are limited to data collected from one research lab and focus exclusively on stroke patients.

**Future work.** To address the complexity in computational notebooks highlighted in the experts’ feedback, in addition to the documentation, following recent works [94], [95], we plan to integrate widgets and buttons into the computational notebooks to further simplify data processing tasks. Furthermore, we plan to create additional training materials, such as video tutorials, to make the framework more easily accessible to researchers and clinicians. We also plan to gather data from various labs and clinics, involving a more diverse group of patients (e.g., cerebral palsy patients, athletes), and include new usage scenarios that demonstrate the framework’s versatility and effectiveness. Also, as future work, we aim to explore new methodologies for developing open-access frameworks designed for widespread adoption, motivated by our prior efforts (e.g., [96], [97]), moving beyond the closed-source bespoke systems typically seen in design studies. While our current work takes a step in this direction by incorporating surveys and interviews with a broad range of experts, we aim to formalize this approach as a distinct methodology in the future.

## ACKNOWLEDGMENTS

We thank the reviewers for their constructive comments and helpful suggestions. Our work is supported by the National Institutes of Health (Award R01-HD088543, Administrative Supplement) and the National Science Foundation (Awards #2320261, #2330565, #2411223).

## REFERENCES

- [1] B. R. Bloem, J. Haan, A. M. Lagaay, W. van Beek, A. R. Wintzen, and R. A. Roos, "Investigation of gait in elderly subjects over 88 years of age," *Journal of Geriatric Psychiatry and Neurology*, vol. 5, no. 2, pp. 78–84, 1992.
- [2] L. Sudarsky, "Gait disorders: prevalence, morbidity, and etiology," *Advances in Neurology*, vol. 87, pp. 111–117, 2001.
- [3] A. Nieuwboer, W. d. Weerd, R. Dom, and E. Lesaffre, "A frequency and correlation analysis of motor deficits in Parkinson patients," *Disability and Rehabilitation*, vol. 20, no. 4, pp. 142–150, 1998.
- [4] J. J. Eng and K. S. Chu, "Reliability and comparison of weight-bearing ability during standing tasks for individuals with chronic stroke," *Archives of Physical Medicine and Rehabilitation*, vol. 83, no. 8, pp. 1138–1144, 2002.
- [5] K. Himmelmann, G. Hagberg, and P. Uvebrant, "The changing panorama of cerebral palsy in Sweden. x. prevalence and origin in the birth-year period 1999–2002," *Acta Paediatrica*, vol. 99, no. 9, pp. 1337–1343, 2010.
- [6] L. E. Mitchell, N. S. Adzick, J. Melchionne, P. S. Pasquariello, L. N. Sutton, and A. S. Whitehead, "Spina bifida," *The Lancet*, vol. 364, no. 9448, pp. 1885–1895, 2004.
- [7] "HAS Motion, Visual3D," [https://wiki.has-motion.com/doku.php?id=visual3d:visual3d\\_overview](https://wiki.has-motion.com/doku.php?id=visual3d:visual3d_overview), [Online; accessed 23-November-2024].
- [8] "Vicon-Nexus," <https://www.vicon.com/software/nexus/>, [Online; accessed 16-July-2024].
- [9] A. Karatsidis, G. Bellusci, H. M. Schepers, M. De Zee, M. S. Andersen, and P. H. Veltink, "Estimation of ground reaction forces and moments during gait using only inertial motion capture," *Sensors*, vol. 17, no. 1, p. 75, 2016.
- [10] D. Hamacher, N. Singh, J. H. Van Dieën, M. Heller, and W. R. Taylor, "Kinematic measures for assessing gait stability in elderly individuals: a systematic review," *Journal of The Royal Society Interface*, vol. 8, no. 65, pp. 1682–1698, 2011.
- [11] G. Vervoort, A. Bengevoord, C. Strouwen, E. M. Bekkers, E. Heremans, W. Vandenberghe, and A. Nieuwboer, "Progression of postural control and gait deficits in Parkinson's disease and freezing of gait: A longitudinal study," *Parkinsonism & Related Disorders*, vol. 28, pp. 73–79, 2016.
- [12] D. M. Mohan, A. H. Khandoker, S. A. Wasti, S. Ismail Ibrahim Ismail Alali, H. F. Jelinek, and K. Khalaf, "Assessment methods of post-stroke gait: A scoping review of technology-driven approaches to gait characterization and analysis," *Frontiers in Neurology*, vol. 12, p. 650024, 2021.
- [13] F. Yang, T. Bhatt, and Y.-C. Pai, "Generalization of treadmill-slip training to prevent a fall following a sudden (novel) slip in over-ground walking," *Journal of Biomechanics*, vol. 46, no. 1, pp. 63–69, 2013.
- [14] Y.-C. Pai, T. Bhatt, F. Yang, E. Wang, and S. Kritchevsky, "Perturbation training can reduce community-dwelling older adults' annual fall risk: a randomized controlled trial," *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, vol. 69, no. 12, pp. 1586–1594, 2014.
- [15] Y. Wang, T. Bhatt, X. Liu, S. Wang, A. Lee, E. Wang, and Y.-C. C. Pai, "Can treadmill-slip perturbation training reduce immediate risk of over-ground-slip induced fall among community-dwelling older adults?" *Journal of Biomechanics*, vol. 84, pp. 58–66, 2019.
- [16] S. Wang, F. Miranda, Y. Wang, R. Rasheed, and T. Bhatt, "Near-fall detection in unexpected slips during over-ground locomotion with body-worn sensors among older adults," *Sensors*, vol. 22, no. 9, 2022.
- [17] M. Burnfield, "Gait analysis: Normal and pathological function," *Journal of Sports Science and Medicine*, vol. 9, no. 2, p. 353, 2010.
- [18] B. H. Dobkin, "Rehabilitation after stroke," *New England Journal of Medicine*, vol. 352, no. 16, pp. 1677–1684, 2005.
- [19] D. H. Sutherland and J. R. Davids, "Common gait abnormalities of the knee in cerebral palsy," *Clinical Orthopaedics and Related Research*, no. 288, pp. 139–147, 1993.
- [20] F. M. Chang, J. T. Rhodes, K. M. Flynn, and J. J. Carollo, "The role of gait analysis in treating gait abnormalities in cerebral palsy," *Orthopedic Clinics*, vol. 41, no. 4, pp. 489–506, 2010.
- [21] M. Pistacchi, M. Gioulis, F. Sanson, E. De Giovannini, G. Filippi, F. Rossetto, and S. Z. Marsala, "Gait analysis and clinical correlations in early Parkinson's disease," *Functional Neurology*, vol. 32, no. 1, p. 28, 2017.
- [22] J. Jonsdottir, M. Recalcati, M. Rabuffetti, A. Casiraghi, S. Boccardi, and M. Ferrarin, "Functional resources to increase gait speed in people with stroke: strategies adopted compared to healthy controls," *Gait & Posture*, vol. 29, no. 3, pp. 355–359, 2009.
- [23] J. Boudarham, N. Roche, D. Pradon, C. Bonnyaud, D. Bensmail, and R. Zory, "Variations in kinematics during clinical gait analysis in stroke patients," *PLoS One*, vol. 8, no. 6, p. e66421, 2013.
- [24] M. Iosa, F. Bini, F. Marinuzzi, A. Fusco, G. Morone, G. Koch, A. Martino Cinnera, S. Bonni, and S. Paolucci, "Stability and harmony of gait in patients with subacute stroke," *Journal of Medical and Biological Engineering*, vol. 36, pp. 635–643, 2016.
- [25] C. Beyaert, R. Vasa, and G. E. Frykberg, "Gait post-stroke: Pathophysiology and rehabilitation strategies," *Neurophysiologie Clinique/Clinical Neurophysiology*, vol. 45, no. 4–5, pp. 335–355, 2015.
- [26] N. G. Moreau, A. W. Bodkin, K. Bjornson, A. Hobbs, M. Soileau, and K. Lahasky, "Effectiveness of rehabilitation interventions to improve gait speed in children with cerebral palsy: systematic review and meta-analysis," *Physical Therapy*, vol. 96, no. 12, pp. 1938–1954, 2016.
- [27] J. Marin, J. J. Marin, T. Blanco, J. de la Torre, I. Salcedo, and E. Martitegui, "Is my patient improving? individualized gait analysis in rehabilitation," *Applied Sciences*, vol. 10, no. 23, p. 8558, 2020.
- [28] M. H. Schwartz, J. P. Trost, and R. A. Wewey, "Measurement and management of errors in quantitative gait data," *Gait & Posture*, vol. 20, no. 2, pp. 196–203, 2004.
- [29] Y. Wu and S. Krishnan, "Statistical analysis of gait rhythm in patients with Parkinson's disease," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 2, pp. 150–158, 2009.
- [30] M. Sangeux and J. Polak, "A simple method to choose the most representative stride and detect outliers," *Gait & Posture*, vol. 41, no. 2, pp. 726–730, 2015.
- [31] N. K. Zakaria, N. M. Tahir, and R. Jailani, "A markerless-based gait analysis and visualization approach for asd children," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, 2021.
- [32] A. Seals, G. Piloni, J. Kim, R. Sanchez, J.-R. Rizzo, L. Charvet, O. Nov, and G. Dove, "'Are they doing better in the clinic or at home?': Understanding clinicians' needs when visualizing wearable sensor data used in remote gait assessments for people with multiple sclerosis," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–16.
- [33] A. R. Anwar, H. Yu, and M. Vassallo, "Gait quantification and visualization for digital healthcare," *Health Policy and Technology*, vol. 9, no. 2, pp. 204–212, 2020.
- [34] M. Boumrah, S. Garbaya, and A. Radgui, "Real-time visual analytics for in-home medical rehabilitation of stroke patient—systematic review," *Medical & Biological Engineering & Computing*, vol. 60, no. 4, pp. 889–906, 2022.
- [35] J.-C. Gonzalez-Islas, O.-A. Dominguez-Ramirez, O. Lopez-Ortega, J. Peña-Ramirez, J.-P. Ordaz-Oliver, and F. Marroquin-Gutierrez, "Crouch gait analysis and visualization based on gait forward and inverse kinematics," *Applied Sciences*, vol. 12, no. 20, p. 10197, 2022.
- [36] M. Mukaino, K. Ohtsuka, H. Tanikawa, F. Matsuda, J. Yamada, N. Itoh, and E. Saitoh, "Clinical-oriented three-dimensional gait analysis method for evaluating gait disorder," *JoVE (Journal of Visualized Experiments)*, no. 133, p. e57063, 2018.
- [37] A. Cantu, M. E. Micó-Amigo, S. Del Din, and S. J. Fernstad, "Parallel assemblies plot, a visualization tool to explore categorical and quantitative data: application to digital mobility outcomes," in *2023 IEEE 16th Pacific Visualization Symposium (PacificVis)*. IEEE, 2023, pp. 21–30.
- [38] J. Bernard, N. Wilhelm, B. Krüger, T. May, T. Schreck, and J. Kohlhammer, "MotionExplorer: Exploratory search in human motion capture data based on hierarchical aggregation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2257–2266, 2013.
- [39] S. Jang, N. Elmqvist, and K. Ramani, "MotionFlow: Visual abstraction and aggregation of sequential patterns in human motion tracking data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 21–30, 2015.
- [40] S. Chen, X. Zhao, B. Luo, and Z. Sun, "Visual browse and exploration in motion capture data with phylogenetic tree of context-aware poses," *Sensors*, vol. 20, no. 18, p. 5224, 2020.



- [41] R. C. Contreras, A. Parnandi, B. G. Coelho, C. Silva, H. Schambra, and L. G. Nonato, "NE-Motion: Visual analysis of stroke patients using motion sensor networks," *Sensors*, vol. 21, no. 13, p. 4482, 2021.
- [42] M. Wagner, D. Slijepcevic, B. Horsak, A. Rind, M. Zeppelzauer, and W. Aigner, "KAVAGait: Knowledge-assisted visual analytics for clinical gait analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 3, pp. 1528–1542, 2018.
- [43] A. Agibetov, K. E. T. Gutierrez, C. E. Catalano, G. Patané, C. Hurschler, and M. Spagnuolo, "GaitViewer: Semantic gait data analysis and visualization tool," in *KDWeb*, 2016.
- [44] A. Rind, D. Slijepčević, M. Zeppelzauer, F. Unglaube, A. Kranzl, and B. Horsak, "Trustworthy visual analytics in clinical gait analysis: A case study for patients with cerebral palsy," in *2022 IEEE Workshop on TRust and Expertise in Visual Analytics (TRES)*. IEEE, 2022, pp. 8–15.
- [45] T. Fujiwara, N. Sakamoto, J. Nonaka, K. Yamamoto, K.-L. Ma *et al.*, "A visual analytics framework for reviewing multivariate time-series data with dimensionality reduction," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 1601–1611, 2020.
- [46] Y. Shi, B. Chen, Y. Chen, Z. Jin, K. Xu, X. Jiao, T. Gao, and N. Cao, "Supporting guided exploratory visual analysis on time series data with reinforcement learning," *IEEE Transactions on Visualization and Computer Graphics*, vol. 30, no. 1, pp. 1172–1182, 2023.
- [47] S. Liu, Y. Tian, Z. Deng, W. Cui, H. Zhang, D. Weng, and Y. Wu, "Relation-driven query of multiple time series," *IEEE Transactions on Visualization and Computer Graphics*, 2024.
- [48] G. Shirato, N. Andrienko, and G. Andrienko, "Exploring and visualizing temporal relations in multivariate time series," *Visual Informatics*, vol. 7, no. 4, pp. 57–72, 2023.
- [49] L. Chen, C. Cheng, H. Wang, X. Wang, Y. Tian, X. Yue, W. Kam-Kwai, H. Zhang, S. Hong, and Q. Li, "FMLens: Towards better scaffolding the process of fund manager selection in fund investments," *IEEE Transactions on Visualization and Computer Graphics*, 2024.
- [50] D. Liu, S. Alnegheimish, A. Zytek, and K. Veeramachaneni, "MTV: Visual analytics for detecting, investigating, and annotating anomalies in multivariate time series," *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. CSCW1, pp. 1–30, 2022.
- [51] L. Stopar, P. Skraba, M. Grobelnik, and D. Mladenec, "StreamStory: Exploring multivariate time series on multiple scales," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 4, pp. 1788–1802, 2018.
- [52] Z. Ye and M. Chen, "Visualizing ensemble predictions of music mood," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 1, pp. 864–874, 2022.
- [53] Z. Deng, S. Chen, T. Schreck, D. Deng, T. Tang, M. Xu, D. Weng, and Y. Wu, "Visualizing large-scale spatial time series with geochron," *IEEE Transactions on Visualization and Computer Graphics*, vol. 30, no. 1, pp. 1194–1204, 2023.
- [54] Y. Yu, D. Kruffy, J. Jiao, T. Becker, and M. Behrisch, "PSEUDO: Interactive pattern search in multivariate time series with locality-sensitive hashing and relevance feedback," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 1, pp. 33–42, 2022.
- [55] Y. Wu, Y. Xu, S. Gao, X. Wang, W. Song, Z. Nie, X. Fan, and Q. Li, "LiveRetro: Visual analytics for strategic retrospect in livestream e-commerce," *IEEE Transactions on Visualization and Computer Graphics*, vol. 30, no. 1, pp. 1117–1127, 2023.
- [56] M. Stein, H. Janetzko, A. Lamprecht, T. Breitkreutz, P. Zimmermann, B. Goldlücke, T. Schreck, G. Andrienko, M. Grossniklaus, and D. A. Keim, "Bring it to the pitch: Combining video and movement data to enhance team sport analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 13–22, 2017.
- [57] H. Zeng, X. Wang, A. Wu, Y. Wang, Q. Li, A. Endert, and H. Qu, "EmoCo: Visual analysis of emotion coherence in presentation videos," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 1, pp. 927–937, 2019.
- [58] Z. Chen, S. Ye, X. Chu, H. Xia, H. Zhang, H. Qu, and Y. Wu, "Augmenting sports videos with viscommentator," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 1, pp. 824–834, 2021.
- [59] J. He, X. Wang, K. K. Wong, X. Huang, C. Chen, Z. Chen, F. Wang, M. Zhu, and H. Qu, "VideoPro: A visual analytics approach for interactive video programming," *IEEE Transactions on Visualization and Computer Graphics*, vol. 30, no. 1, pp. 87–97, 2023.
- [60] K. K. Wong, X. Wang, Y. Wang, J. He, R. Zhang, and H. Qu, "Anchorage: Visual analysis of satisfaction in customer service videos via anchor events," *IEEE Transactions on Visualization and Computer Graphics*, vol. 30, no. 7, pp. 4008–4022, 2024.
- [61] P. C. Dixon, J. J. Loh, Y. Michaud-Paquette, and D. J. Pearsall, "biomechZoo: An open-source toolbox for the processing, analysis, and visualization of biomechanical movement data," *Computer Methods and Programs in Biomedicine*, vol. 140, pp. 1–10, 2017.
- [62] "Kinovea," <https://www.kinovea.org/>, [Online; accessed 16-July-2024].
- [63] S. D. Uhrlich, A. Falisse, Ł. Kidziński, J. Muccini, M. Ko, A. S. Chaudhari, J. L. Hicks, and S. L. Delp, "OpenCap: Human movement dynamics from smartphone videos," *PLoS Computational Biology*, vol. 19, no. 10, p. e1011462, 2023.
- [64] Ł. Kidziński, B. Yang, J. L. Hicks, A. Rajagopal, S. L. Delp, and M. H. Schwartz, "Deep neural networks enable quantitative movement analysis using single-camera videos," *Nature Communications*, vol. 11, no. 1, p. 4054, 2020.
- [65] "TEMPLO for Clinical Gait Analysis," <https://contemplas.com/technologie/>, [Online; accessed 16-July-2024].
- [66] "LEGsYS and BalanSens, BioSensics," <https://biosensics.com/technologies/gait-and-balance>, [Online; accessed 16-July-2024].
- [67] B. Bilney, M. Morris, and K. Webster, "Concurrent related validity of the GAITRite walkway system for quantification of the spatial and temporal parameters of gait," *Gait & Posture*, vol. 17, no. 1, pp. 68–74, 2003.
- [68] D. J.-L. Lee, D. Tang, K. Agarwal, T. Boonmark, C. Chen, J. Kang, U. Mukhopadhyay, J. Song, M. Yong, M. A. Hearst *et al.*, "Lux: always-on visualization recommendations for exploratory dataframe workflows," *Proceedings of the VLDB Endowment*, vol. 15, no. 3, pp. 727–738, 2021.
- [69] A. Tritsarolis, C. Doulkeridis, N. Pelekis, and Y. Theodoridis, "St\_visions: a Python library for interactive visualization of spatio-temporal data," in *2021 22nd IEEE International Conference on Mobile Data Management (MDM)*. IEEE, 2021, pp. 244–247.
- [70] Y. Chen and E. Wu, "Pi2: End-to-end interactive visualization interface generation from queries," in *Proceedings of the 2022 International Conference on Management of Data*, 2022, pp. 1711–1725.
- [71] X. Li, Y. Zhang, J. Leung, C. Sun, and J. Zhao, "Edassistant: Supporting exploratory data analysis in computational notebooks with in situ code search and recommendation," *ACM Transactions on Interactive Intelligent Systems*, vol. 13, no. 1, pp. 1–27, 2023.
- [72] W. Epperson, V. Gorantla, D. Moritz, and A. Perer, "Dead or alive: Continuous data profiling for interactive data science," *IEEE Transactions on Visualization and Computer Graphics*, vol. 30, no. 1, pp. 197–207, 2023.
- [73] J. Palmeiro, B. Malveiro, R. Costa, D. Polido, R. Moreira, and P. Bizarro, "Data+Shift: Supporting Visual Investigation of Data Distribution Shifts by Data Scientists," in *EuroVis 2022 - Short Papers*. The Eurographics Association, 2022.
- [74] Z. J. Wang, A. Kale, H. Nori, P. Stella, M. E. Nunnally, D. H. Chau, M. Vorvoreanu, J. Wortman Vaughan, and R. Caruana, "Interpretability, then what? editing machine learning models to reflect human knowledge and values," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 4132–4142.
- [75] P. Xenopoulos, J. Rulff, L. G. Nonato, B. Barr, and C. Silva, "Calibrate: Interactive analysis of probabilistic model output," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 1, pp. 853–863, 2022.
- [76] G. Guo, E. Karavani, A. Endert, and B. C. Kwon, "Causalvis: Visualizations for causal inference," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–20.
- [77] G. Guo, M. Glenski, Z. Shaw, E. Saldanha, A. Endert, S. Volkova, and D. Arendt, "Vaine: Visualization and AI for natural experiments," in *2021 IEEE Visualization Conference (VIS)*. IEEE, 2021, pp. 21–25.
- [78] R. Li, W. Cui, T. Song, X. Xie, R. Ding, Y. Wang, H. Zhang, H. Zhou, and Y. Wu, "Causality-based visual analysis of questionnaire responses," *IEEE Transactions on Visualization and Computer Graphics*, vol. 30, no. 1, pp. 638–648, 2023.
- [79] X. Teng, Y. Ahn, and Y.-R. Lin, "VISPUR: Visual aids for identifying and interpreting spurious associations in data-driven decisions," *IEEE Transactions on Visualization and Computer Graphics*, vol. 30, no. 1, pp. 219–229, 2023.
- [80] Z. Jiang, H. Chen, R. Zhou, J. Deng, X. Zhang, R. Zhao, C. Xie, Y. Wang, and E. C. Ngai, "HealthPrism: a visual analytics system for exploring children's physical and mental health profiles with multimodal data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 30, no. 1, pp. 1205–1215, 2023.
- [81] H. Yang, Y. Li, and M. X. Zhou, "Understand users' comprehension and preferences for composing information visualizations," *ACM Transactions on Computer-Human Interaction*, vol. 21, no. 1, pp. 1–30, 2014.
- [82] G. J. Quadri, A. Z. Wang, Z. Wang, J. Adorno, P. Rosen, and D. A. Szafir, "Do you see what I see? a qualitative study eliciting high-level

- visualization comprehension,” in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–26.
- [83] Q. Zhu, T. Lu, S. Guo, X. Ma, and Y. Yang, “CompositingVis: Exploring interactions for creating composite visualizations in immersive environments,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 31, no. 1, pp. 591–601, 2025.
  - [84] K. Jansen, F. De Groote, F. Massaad, P. Meyns, J. Duysens, and I. Jonkers, “Similar muscles contribute to horizontal and vertical acceleration of center of mass in forward and backward walking: implications for neural control,” *Journal of Neurophysiology*, vol. 107, no. 12, pp. 3385–3396, 2012.
  - [85] M. Serrao, M. Rinaldi, A. Ranavolo, F. Lacquaniti, G. Martino, L. Leonardi, C. Conte, T. Varrecchia, F. Draicchio, G. Coppola *et al.*, “Gait patterns in patients with hereditary spastic paraparesis,” *PLoS One*, vol. 11, no. 10, p. e0164623, 2016.
  - [86] S. Blau, L. Davis, A. Gorney, C. Dohse, K. Williams, J. Lim, W. Pfitzner, E. Laber, G. Sawicki, and N. Olby, “Quantifying center of pressure variability in chondrodystrophoid dogs,” *The Veterinary Journal*, vol. 226, pp. 26–31, 2017.
  - [87] E. Troisi Lopez, R. Minino, P. Sorrentino, R. Rucco, A. Carotenuto, V. Agosti, D. Tafuri, V. Manzo, M. Liparoti, and G. Sorrentino, “A synthetic kinematic index of trunk displacement conveying the overall motor condition in Parkinson’s disease,” *Scientific Reports*, vol. 11, no. 1, p. 2736, 2021.
  - [88] B. C. Kwon, M.-J. Choi, J. T. Kim, E. Choi, Y. B. Kim, S. Kwon, J. Sun, and J. Choo, “RetainVis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 299–309, 2018.
  - [89] C. D. Linhares, D. M. Lima, J. R. Ponciano, M. M. Olivatto, M. A. Gutierrez, J. Poco, C. Traina, and A. J. Traina, “ClinicalPath: A visualization tool to improve the evaluation of electronic health records in clinical decision-making,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 10, pp. 4031–4046, 2022.
  - [90] A. Wentzel, S. Attia, X. Zhang, G. Canahuat, C. D. Fuller, and G. E. Marai, “DITTO: A visual digital twin for interventions and temporal treatment outcomes in head and neck cancer,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 31, no. 1, pp. 65–75, 2025.
  - [91] A. Endert, M. S. Hossain, N. Ramakrishnan, C. North, P. Fiaux, and C. Andrews, “The human is the loop: new directions for visual analytics,” *Journal of Intelligent Information Systems*, vol. 43, pp. 411–435, 2014.
  - [92] T. Munzner, *Visualization analysis and design*. CRC press, 2014.
  - [93] M. Bostock, V. Ogievetsky, and J. Heer, “D<sup>3</sup> data-driven documents,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2301–2309, 2011.
  - [94] M. B. Kery, D. Ren, F. Hohman, D. Moritz, K. Wongsuphasawat, and K. Patel, “mage: Fluid moves between code and graphical work in computational notebooks,” in *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, 2020, pp. 140–151.
  - [95] F. Choi, S. Rahman, H. Kim, and D. Zhang, “Towards transparent, reusable, and customizable data science in computational notebooks,” in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–8.
  - [96] G. Moreira, M. Hosseini, M. N. Alam Nipu, M. Lage, N. Ferreira, and F. Miranda, “The Urban Toolkit: A grammar-based framework for urban visual analytics,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 30, no. 1, pp. 1402–1412, 2024.
  - [97] G. Moreira, M. Hosseini, C. Veiga, L. Alexandre, N. Colaninno, D. de Oliveira, N. Ferreira, M. Lage, and F. Miranda, “Curio: A dataflow-based framework for collaborative urban visual analytics,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 31, no. 1, pp. 1224–1234, 2025.

**Shuaijie Wang** is a Research Associate Professor in the Cognitive, Motor and Balance Rehabilitation Laboratory at the University of Illinois Chicago. He earned his Ph.D. from the University of Chinese Academy of Sciences. His expertise lies in biomedical engineering, with a focus on gait analysis, balance control, fall prevention, and fall risk assessment. His research involves understanding the mechanisms of falls, developing training programs to reduce fall risk in older adults, and using machine learning to model and assess fall risk.

**Ridhuparan Kungumaraju** is a Master’s student in Computer Science at the University of Illinois Chicago. His research interests include visual data science and software development.

**Tanvi Bhatt** is a Professor in the Department of Applied Health Sciences at the University of Illinois Chicago. She is the Director of the Cognitive, Motor and Balance Rehabilitation Laboratory and Co-Director of the Clinical Gait and Motion Analysis Laboratory. She received her Ph.D. in Movement Sciences/Motor Control from the University of Illinois at Chicago. Her research focuses on adaptive perturbation training for fall prevention, investigating the neuromechanical basis of balance recovery from slips and trips, and designing interventions to reduce fall risk.

**Fabio Miranda** is an Assistant Professor in the Department of Computer Science at the University of Illinois Chicago. He received his Ph.D. in Computer Science from New York University. His research focuses on large-scale data analysis, data structures, and data visualization, and visual analytics.

**Kazi Shahrukh Omar** is Ph.D. student in Computer Science at the University of Illinois Chicago. His research interests include visualization and visual analytics, urban computing, and health informatics.