Predicting Multiple Sclerosis From Gait Dynamics Using an Instrumented Treadmill: A Machine Learning Approach

Rachneet Kaur, Zizhang Chen, Robert Motl, Manuel E. Hernandez, *Member, IEEE*, and Richard Sowers, *Member, IEEE*

Abstract*—*Objective: **Multiple Sclerosis (MS) is a neurological condition which widely affects people 50-60 years of age. While clinical presentations of MS are highly heterogeneous, mobility limitations are one of the most frequent symptoms. This study examines a machine learning (ML) framework for identifying MS through spatiotemporal and kinetic gait features.** ***Methods:*** **In this study, gait data during self-paced walking on an instrumented treadmill from 20 persons with MS and 20 age, weight, height, and gender-matched healthy older adults (HOA) were obtained. We explored two strategies to normalize data and minimize dependence on subject demographics;** ***size-normalization*** **(standard body size-based normalization) and** ***regress-normalization*** **(regression-based normalization using scaling factors derived by regressing gait features on multiple subject demographics); and proposed an ML based methodology to classify individual strides of older persons with MS (PwMS) from healthy controls. We generalized both across different walking tasks and subjects.** ***Results:*** **We observed that** ***regress-normalization*** **improved the accuracy of identifying pathological gait using ML when compared to** ***size-normalization*****. When generalizing from comfortable walking to walking while talking, gradient boosting machine achieved the optimal subject classification accuracy and AUC of 94.3 and 1.0, respectively and for subject generalization, a multilayer perceptron resulted in the best accuracy and AUC of 80% and 0.86, respectively, both with regression-normalized data.** ***Conclusion:*** **The integration of gait data and ML may provide a viable patient-centric approach to aid clinicians in monitoring MS.** ***Significance:*** **The results of this study have future implications for the way regression normalized gait features may be clinically used to design ML-based disease prediction strategies and monitor disease progression in PwMS.**

***Index Terms*—Multiple sclerosis, gait, machine learning, conditional entropy, progression space.**

I. Introduction

MULTIPLE Sclerosis (MS) is a chronic demyelinating and neurodegenerative disorder that impairs the central nervous system. It can affect a range of cognitive, physical, and psychiatric processes [1], [2]. Severe symptoms include impairment of vision and sensory abilities, muscle paralysis, and depression [3], with mobility impairments being one of the most frequent signs [4]. MS affects approximately 1 million people in the United States (US) and more than 2 million globally [5]. Peak prevalence is in adults 50-60 years of age [6]. Direct medical treatment expenses and indirect costs in terms of lost productivity, additional need for caretakers and amenities for persons with MS (PwMS) are estimated to be $24 billion annually in the US [7].

Walking and balance difficulties are one of the most common indicators in PwMS; nearly 85% of PwMS describe gait disorders as a major complication [8] and roughly 50% patients need walking assistance within 15 years of MS onset [9]. Secondary effects often include fear of falling, significantly impacting the quality of life of PwMS [10]. In contrast to the monitoring of most underlying manifestations of MS, which require neurological examinations by a trained practitioner, gait can be quickly and remotely monitored. Thus, objective gait monitoring, which expands upon typical clinical tests [11], may be important for designing disease prediction and progression strategies in PwMS. Past research on MS assessment with gait-related dynamics has typically relied upon statistical inferences that may not be able to gauge the heterogeneity present in the disease [12]–[16]. Given that subtle and heterogeneous patterns of gait changes may arise in PwMS over time, a machine learning (ML) approach will be valuable for monitoring MS-related changes in older adults.

This study aims to examine MS and disability related changes in spatiotemporal and kinetic gait features after normalization; and evaluate the effectiveness of a **g**ait data-based machine learning (**ML**) framework **for MS** prediction (**GML4MS**), an

Manuscript received October 5, 2020; revised November 29, 2020; accepted December 22, 2020. Date of publication December 30, 2020; date of current version August 20, 2021. This work was supported in part by the University of Illinois Center for Wearable Intelligent Technologies SRI. The protocol for this study was approved under the University of Illinois at Urbana-Champaign Institutional Review Board number 15674 on 4/3/2015. *(Corresponding author: Rachneet Kaur.)*

Rachneet Kaur is with the Department of Industrial, and Enterprise Systems Engineering at the University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: rk4@illinois.edu).

Zizhang Chen is with the Department of Mathematics and the Department of Statistics at the University of Illinois at Urbana-Champaign, USA.

Robert Motl is with the Department of Physical Therapy, University of Alabama at Birmingham, USA.

Manuel E. Hernandez is with the Department of Kinesiology and Community Health, University of Illinois at Urbana-Champaign, USA.

Richard Sowers is with the Department of Industrial and Enterprise Systems Engineering and the Department of Mathematics, University of Illinois at Urbana-Champaign, USA.

This article has supplementary downloadable material available at https://doi.org/10.1109/TBME.2020.3048142, provided by the authors.

Digital Object Identifier 10.1109/TBME.2020.3048142

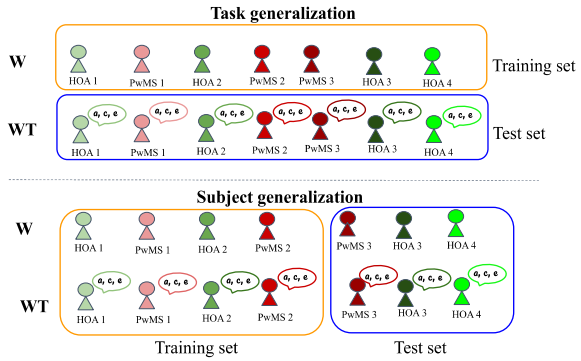

Fig. 1. Top: Task generalization model, Bottom: Subject generalization design. Healthy older adults (HOA) and PwMS are depicted in shades of green and red, respectively. The indices $(1, 2, 3, \ldots)$ along with HOA and PwMS are used as a reference for dummy subject identifies.

ML-based methodology to classify strides of older PwMS from healthy controls, so as to generalize across different walking tasks and subjects after gait normalization. Building upon prior work examining MS-related variations in gait characteristics [17], we categorized PwMS using the following two classification designs (see Fig. 1):

a) Task generalization establishing the generality over different tasks. In these tasks, we train binary (healthy vs. MS) classifiers on walking (W) trials and apply them to walking while talking (WT) trials. Task generalization results will hopefully reflect how classifiers trained in supervised lab conditions might work in real-world gait tasks with challenges of divided attention. To monitor disease progression and relapses, task generality is vital as normative data collected in a clinic or lab could be used as a basis to assess gait data collected using wearable sensors in a home-based setting.

b) Subject generalization demonstrating the generality over different subjects. In these tasks, we train binary (healthy vs. MS) classifiers (with a balanced collection of W and WT tasks) on some test subjects and apply them to the withheld separate set of test subjects. These results may have implications in detection of MS in new patients.

II. Related Results and Context

Several studies have identified gait performance declines in PwMS, particularly as disability increases [12], [18], [19]. Most gait-based methods for identifying MS have relied upon traditional statistical techniques to examine differences in spatiotemporal features and correlations with disability [12]–[16]. Supervised ML methodologies such as random forest and artificial neural networks have already been used in human gait analysis across other neurological populations [20], [21]. A few prior works have explored ML to classify MS using gait data [22], [23]. However, to the best of our knowledge, there is no study utilizing ML on spatiotemporal and kinetic gait characteristics for MS prediction. Despite model-based statistical practices presenting transparency and explainability regarding the contribution of independent features, ML approaches may improve performance by addressing high-dimensional and non-linear feature interactions in a model-free way. Further, transforming statistical inference to prediction classes requires defining sensitive classification thresholds.

Distinctive physical characteristics across subjects inherently enhance the variability in raw gait parameters and thus limit the efficiency of learning true reliable trends in a feature differentiating healthy and pathological gait [24]. Referring to the performance improvements in previous studies examining neurological diagnosis [21], [25], two normalization strategies, namely *size-N* (standard body size-based normalization) and *regress-N* (regression-based normalization using scaling factors derived by regressing gait features on multiple subject demographics) were explored to minimize the dependency of gait features on the subject demographics.

The proposed application of ML classifiers to recognize gait patterns of PwMS across tasks and over new subjects is a step forward towards the identification of a tipping point for older people with MS, and worsening of symptoms in the near term. Moreover, we discuss the importance of spatiotemporal and kinetic features, encompassing valuable domain knowledge, in the classification performance. Attributing to prior evidence of gait changes with MS impairment [12], [26], [27], we construct an *MS progression space* by unsupervised clustering of *reduced* gait feature space in PwMS to examine the relative correspondence of the defined subgroups to disease severity. This analysis may facilitate strategies to monitor disease progression and evaluate the effectiveness of disease modifying interventions. The proposed methodology is an advancement towards developing an assessment marker for medical professionals to predict older PwMS who are likely to have a worsening of symptoms in the near term. Our ultimate objective is a system to automatically identify inflection points in the disease progression of older PwMS.

III. Experimental Design: Setup and Subjects

A. Experimental Paradigm

An instrumented treadmill (C-Mill, Motekforce Link, Culemborg, The Netherlands) in self-paced mode was utilized to allow subjects to walk at their preferred speed. To allow for unbiased force recordings, subjects were instructed to restrain from holding the handrails while walking on the treadmill. For safety purposes, all subjects wore a ceiling-mounted harness and had access to an emergency stop button during all the walking trials. Supplementary figure S1 illustrates the gait data acquisition setup. All subjects walked one trial under two different task conditions, namely single-task condition, W and dual-task paradigm, WT. For the WT task, subjects were asked to walk while reciting alternate letters of the alphabet (i.e. a, c, e, $\ldots$), coordinating equal attention between mobility and the cognitive interference exercise to depreciate the influence of task prioritization. The divided attention dual-task walking in a laboratory environment has been demonstrated to be more analogous (as compared to usual walking) to every-day walking in the older adults and hence provides a competent framework to generalize adequacy towards daily-living gait for 24/7 monitoring scenarios [28]. Further, the attention demanding WT task has been examined by researchers for practical implications in designing mobility risk assessment procedures and predicting the risk of falls and fall-related injuries in older adults and individuals with other cognitive or movement disorders [29]. For each trial, subjects were instructed to walk at a comfortable pace for up to 75 seconds (s), after being provided with a brief

training session. CueFors 2 software [30] was used to collect gait event data (i.e., left and right heel strike, mid-stance, and toe-off position coordinates and time stamps) and raw data (i.e., vertical ground reaction forces, treadmill speed and center of pressure (CoP) position coordinates at a 500 Hz frequency) during each walking trial. To facilitate the online identification of gait events, an online pattern recognition algorithm detects maxima and minima in the butterfly patterns (see Section IV-B4) of the CoP profiles, that are collected in real time via an embedded force plate in the treadmill [31]. Supplementary table S2 describes the collected raw features.

B. Study Participants

Twenty individuals from each cohort, MS patients (age: 61.05 ± 6.87 years $[49 - 75$ years$]$, weight: 74.89 ± 24.52 kg $[21.6 - 135$ kg$]$, height: 1.68 ± 0.09 m $[1.6 - 1.93$ m$]$, male/female: $5/15$) and healthy older adults (HOA) (age: 61.2 ± 5.87 years $[48 - 68$ years$]$, weight: 76.17 ± 19.24 kg $[52.1 - 121$ kg$]$, height: 1.70 ± 0.07 m $[1.56 - 1.90$ m$]$, male/female: $5/15$) were recruited from the local community. All subjects were medically stable, right-side dominant, had no lower limb injury in the past six months and had normal or corrected to normal vision. MS subjects had mild to moderate disability (4.3 ± 1.62 $[1.0 - 6.0]$ as evaluated by the Kurtzke's Expanded Disability Status Scale (EDSS) [32]), were relapse-free for at least a month prior to experimental trials and had no other cognitive dysfunction or neurological disorders. EDSS, monitoring sensory, motor, brain stem, visual, cerebellar, bowel and bladder, pyramidal and other functions, is an accepted method to quantify disability in PwMS. For this work, we divided PwMS into three sub groups based on their EDSS score: mild (1.0–2.5), mild-to-moderate (3.0–4.5) and moderate (5.0–6.0). No significant differences (at significance level $\alpha = 0.05$) in age, weight, height, gender and education levels were observed between the two cohorts. Two HOA and three PwMS were excluded from the analysis for holding the handrails (biasing the raw force data).

IV. Experimental Design: Data Analysis

A. Gait Terminology and Mathematical Notation

A typical walking gait comprises of recurrent gait cycles (GC). A gait cycle or stride is measured from a foot's heel strike to the subsequent heel strike of the same foot. For our analysis, a stride was characterized by the following gait events: HSR: heel strike right, TOL: toe-off left, MidSSR: midstance right, HSL: heel strike left, TOR: toe-off right, MidSSL: midstance left, with the next HSR starting a new stride. A stride is a consolidation of two steps (i.e. HSR-HSL and HSL-HSR), where a step is marked from a foot's heel strike to the following heel strike of the opposite foot. Supplementary figure S3 demonstrates the longitudinal plane view of a GC. The following are frequently used mathematical notations:

- Let N_s be the total number of valid strides recorded during a subject's complete walking trial on the treadmill
- Let $(\tilde{S}, \leq_s)$ where $\tilde{S} \stackrel{\text{def}}{=} \{s_k, k = 1, 2, \ldots, N_s\}$ be an ordered set of valid strides during the complete walk where $s_m \leq_s s_n$ essentially means that stride s_m appeared prior in the subject's walk to stride s_n. Since the strides derived from a trial are ordered in time, $s_m \leq_s s_n$ if $m \leq n$ defines a natural order on $\tilde{S}$. Clearly, cardinality $|\tilde{S}| = N_s$.
- Let $(E, \lhd)$ be an ordered set of six gait events observed during a stride

$$E \stackrel{\text{def}}{=} \{HSR, TOL, MidSSR, HSL, TOR, MidSSL\}$$

where the order $\lhd$ is defined as follows:

$$HSR \lhd TOL \lhd MidSSR \lhd HSL \lhd TOR \lhd MidSSL$$

- Let $T_{\text{raw}} \stackrel{\text{def}}{=} \left\{\delta t,\ t = 0, 1, 2, \ldots, \frac{T_{\text{walk}}}{0.002}\right\}$ be the times (in s) corresponding to raw force and CoP recordings where $\delta = 0.002$, $T_{walk} = 75$ since the raw data is collected every 0.002 s and each trial lasted for 75 s. For each time stamp $t \in T_{\text{raw}}$, define:
  - $S(t)$ as the treadmill speed (in m/s)
  - $F_Z(t)$ as the ground reaction force (in Newton (N))
  - $(CoPX(t), CoPY(t))$ as the CoP positions in x and y-directions (in m)
- Define the Cartesian product $(E \times \tilde{S}, \prec)$ where $E \times \tilde{S} = \{(e, s_k) : e \in E \text{ and } s_k \in \tilde{S}\}$ as the set of ordered pairs (e, s_k) corresponding to event e of stride s_k for every $e \in E$ and $s_k \in \tilde{S}$ where eq. (1) defines the ordering on $E \times \tilde{S}$.

$$(e, s_m) \prec (f, s_n) \text{ if } \begin{cases} s_m <_s s_n & \text{for } m \neq n \\ e \lhd f & \text{for } m = n \end{cases} \quad (1)$$

For each gait event and stride $(e, s_k) \in E \times \tilde{S}$, define:

- $T_e^{(s_k)}$ as the elapsed time (in s) from the start of data recording to (e, s_k);
- $\left(X_e^{(s_k)}, Y_e^{(s_k)}\right)$ as the x and y-coordinates (relative to origin of the treadmill) for the detected (e, s_k);
- $\tilde{T}_e^{(s_k)} \stackrel{\text{def}}{=} \min\{t : t > T_e^{(s_k)} \text{ and } t \in T_{\text{raw}}\}$ as the closest time in T_{raw} (corresponding to the recorded raw forces and CoP positions) to the marked time $T_e^{(s_k)}$;
- $F_Z^{(e,s_k)} \stackrel{\text{def}}{=} F_Z(\tilde{T}_e^{(s_k)})$ as the reaction force at (e, s_k);
- $\widehat{CoP}_{e,f}^{(s_k),(s_m)}$ as the CoP trajectory between $(e, s_k), (f, s_m) \in E \times \tilde{S}$ (events e and f of strides s_k and s_m, respectively) where $(e, s_k) \prec (f, s_m)$ eq. (1)

$$\widehat{CoP}_{e,f}^{(s_k),(s_m)} \stackrel{\text{def}}{=} \{(CoPX(t), CoPY(t)) : \tilde{T}_e^{(s_k)} \leq t \leq \tilde{T}_f^{(s_m)}\}$$

B. Gait Feature Extraction for MS Characterization

To examine cohort related variations in the gait patterns, characteristic kinematic and kinetic features were extracted across strides from the raw gait data using Python 3.6 (see supplementary figure S4 for our workflow pipeline). The derived features can be categorized as follows:

1) Temporal Features: 7 temporal gait features, namely stride time, stance time, swing time, supporting (right single, initial double and terminal double) times (in s) and cadence (in $steps/min$) were computed for each stride.

- Stride time is the time between two successive heel strikes of the same foot i.e. HSR-HSR. ST (2) denotes the set of stride times for a complete trial.

$$ST = \{ST^{(s_k)} : s_k \in \tilde{S}\} \text{ where } ST^{(s_k)} = T_{HSR}^{(s_{k+1})} - T_{HSR}^{(s_k)} \quad (2)$$

- Stance time $(S_tT^{(s_k)} = T_{TOR}^{(s_k)} - T_{HSR}^{(s_k)})$ is the time between heel strike and toe-off (from stride $s_k \in \tilde{S}$) of the same foot i.e. HSR-TOR.
- Swing time $(S_wT^{(s_k)} = T_{HSR}^{(s_{k+1})} - T_{TOR}^{(s_k)})$ is measured between the toe-off (TOR, s_k) and heel strike (HSR, s_{k+1}) of the same foot.
- Support can be categorized as single or double depending on whether only one or both of the subject's feet are in contact with the treadmill's belt, respectively. Single support can further be classified as left/right depending on which one foot supports the subject's body.
  - Left single supporting time $(SS_L^{(s_k)} = T_{HSR}^{(s_{k+1})} - T_{TOR}^{(s_k)})$ is the time between toe-off (TOR, s_k) and heel strike (HSR, s_{k+1}) of the right foot for stride $s_k \in \tilde{S}$. This is identical to swing time.
  - Right single supporting time $(SS_R^{(s_k)} = T_{HSL}^{(s_k)} - T_{TOL}^{(s_k)})$ is the time between toe-off (TOL, s_k) and heel strike (HSL, s_k) of the left foot for stride $s_k \in \tilde{S}$.

  Double support can be identified as initial/terminal based on it's onset in the stance phase.
  - Initial double supporting time $(DS_I^{(s_k)} = T_{TOL}^{(s_k)} - T_{HSR}^{(s_k)})$ is the time amid heel strike of supporting foot and toe-off of other foot i.e. HSR-TOL from stride $s_k \in \tilde{S}$.
  - Terminal double supporting time $(DS_T^{(s_k)})$ is calculated between heel strike of the other foot and toe-off of the supporting foot i.e. HSL-TOR from stride s_k.

$$DS_T = \{DS_T^{(s_k)} : s_k \in \tilde{S}\} \text{ where } DS_T^{(s_k)} = T_{TOR}^{(s_k)} - T_{HSL}^{(s_k)}$$

- Cadence $(C^{(s_k)} = 60 \times 2/(T_{HSR}^{(s_{k+1})} - T_{HSR}^{(s_k)}))$ is the walking rate or number of steps taken in a minute (min) i.e. twice the inverse of stride time (in min) for stride $s_k \in \tilde{S}$.

2) Spatial Features: The stride-wise extracted 4 spatial (distance dimension) gait attributes included stride length, stride width (in m) and the dimensionless left and right foot progression angles. Since the foot comes back to its initial position after each stride while walking on a treadmill belt, the y-coordinate of position for the current and next stride event, HSR for instance, will be approximately the same each time. Therefore, to report accurate spatial measures, y-position coordinates were corrected to account for the relative treadmill belt travel (BT). Mathematically, $BT((e, s_m), (f, s_n)) = \int_{t_1=\tilde{T}_e^{(s_m)}}^{t_2=\tilde{T}_f^{(s_n)}} S(t)dt$ is computed as the area under the speed-time curve bounded by the closest times (corresponding to recorded speeds) to the marked times of gait events $(e,\ s_m)$ and $(f,\ s_n) \in E \times \tilde{S}$ where $(e,\ s_m) \prec (f,\ s_n)$ and $dt = 0.002$. The above integral is numerically approximated via the trapezoidal rule. Hence, the relative y-coordinate for (f, s_n) w.r.t (e, s_m) is given by (3).

$$\widehat{Y}_f^{(s_n)} = Y_f^{(s_n)} + BT((e, s_m), (f, s_n)) \quad (3)$$

Now, let's define the derived spatial gait markers.

- Stride length $(SL^{(s_k)})$ is the horizontal distance in the walking plane between two subsequent heel strikes of the same foot i.e. between (HSR, s_k) and (HSR, s_{k+1}).

$$SL = \{SL^{(s_k)} : s_k \in \tilde{S}\} \text{ where } SL^{(s_k)} = \widehat{Y}_{HSR}^{(s_{k+1})} - Y_{HSR}^{(s_k)}$$

  where $\widehat{Y}$ are adjusted for belt travel relative to (HSR, s_k).
- Stride width $(SW^{(s_k)})$ is the medio-lateral distance between the two feet i.e. perpendicular distance between the line connecting two consecutive heel strikes of the same foot i.e. (HSR, s_k) and (HSR, s_{k+1}) and the heel strike of the contralateral foot i.e. (HSL, s_k).

$$SW^{(s_k)} = \frac{1}{D^{(s_k)}} \left| \left(X_{HSR}^{(s_{k+1})} - X_{HSR}^{(s_k)}\right)\left(Y_{HSR}^{(s_k)} - \widehat{Y}_{HSL}^{(s_k)}\right) - \left(X_{HSR}^{(s_k)} - X_{HSL}^{(s_k)}\right)\left(\widehat{Y}_{HSR}^{(s_{k+1})} - Y_{HSR}^{(s_k)}\right) \right|$$

  where
  $D^{(s_k)} = \sqrt{(X_{HSR}^{(s_{k+1})} - X_{HSR}^{(s_k)})^2 + (\widehat{Y}_{HSR}^{(s_{k+1})} - Y_{HSR}^{(s_k)})^2}$
  and $\widehat{Y}$ are adjusted for belt travel relative to (HSR, s_k).
- Foot progression angle (FPA) for the right/left $(\theta_R^{(s_k)}/\theta_L^{(s_k)})$ foot is defined as the angle between the progression vector (P_R/P_L) (joining two consecutive heel strikes of the right/left foot) and the foot vector (F_R/F_L) (drawn between the right/left foot's heel strike and toe-off) for stride s_k [33]. Since staggered walking in PwMS might show significant fluctuations in FPAs, we elected it as a potential feature correlating to MS gait. Mathematically, we have:

$$\theta_* = \left\{ \theta_*^{(s_k)} = (-1)^x \tan^{-1}\left(\frac{Y_{P_*}^{(s_k)}}{X_{P_*}^{(s_k)}}\right) + \cdots (-1)^y \tan^{-1}\left(\frac{Y_{F_*}^{(s_k)}}{X_{F_*}^{(s_k)}}\right) : s_k \in \tilde{S} \right\}$$

$$(X_{F_*}^{(s_k)},\ Y_{F_*}^{(s_k)}) = (X_{TO*}^{(s_{k_1})} - X_{HS*}^{(s_k)},\ \widehat{Y}_{TO*}^{(s_{k_1})} - Y_{HS*}^{(s_k)})$$

$$(X_{P_*}^{(s_k)},\ Y_{P_*}^{(s_k)}) = (X_{HS*}^{(s_k)} - X_{HS*}^{(s_{k-1})},\ \widehat{Y}_{HS*}^{(s_k)} - Y_{HS*}^{(s_{k-1})})$$

  where $\widehat{Y}$ are adjusted y-coordinates relative to the belt travel (3), $*$ indicates left (L) or right (R) and s_{k_1} denotes s_{k+1} and s_k for L and R, respectively. Exponents x, y are defined as 1, 2 respectively for L and 2, 1 respectively for R. Supplementary figure S5 summarizes these definitions on an overground view of the GCs.

3) Spatiotemporal Features: Derived from the above defined temporal and spatial features, 2 additional spatiotemporal markers, namely stride speed (in m/s) and walk ratio (in m/$strides/min$) were defined for each GC.

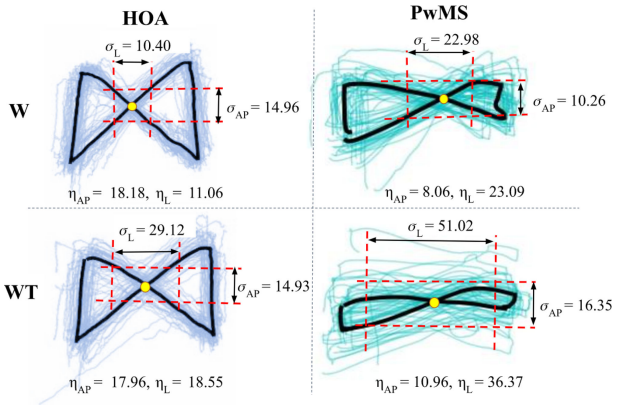

Fig. 2. Butterfly diagram. Left: HOA, Right: PwMS with EDSS = 5.5, Top: Trial W, Bottom: Trial WT. The curves illustrate the BD during the entire 75 s walk where the thicker black line and yellow circle depicts the mean trajectory and mean IP, respectively. Here, x and y axis represent the CoP position coordinates. The variability (red dashed lines) and asymmetry in the AP (σ_{AP}/η_{AP}) and lateral (σ_L/η_L) directions are reported in mm.

- Stride speed ($SS^{(s_k)} = SL^{(s_k)}/ST^{(s_k)}$) is defined as the ratio of stride length and stride time for strides $s_k \in \tilde{S}$.
- Walk ratio ($W^{(s_k)} = 2 \times SL^{(s_k)}/C^{(s_k)}$) is computed as the ratio of stride length to the number of strides walked per minute (i.e. half the cadence) for GCs $s_k \in \tilde{S}$.

4) Kinetic Features: 8 kinetic gait parameters, namely the six forces, one at each gait event (in N) and two butterfly diagram-based features (in m) were identified for each GC.

- Forces ($F_Z^{(e,s_k)}$) at each of the six gait events ($e \in E$) were recorded for every stride s_k. Thus, for a trial, we have

$$F_Z^e = \left\{ F_Z^{(e,s_k)} : (e, s_k) \in \{e\} \times \tilde{S} \right\} \; \forall \, e \in E$$

- Butterfly diagram (BD) reflects the repeated CoP trajectory for multiple continuous strides during a subject's walk. The BD derived features, especially in the anterior-posterior (AP) and lateral directions, have been associated with important neurological functions in PwMS [34] (Fig. 2). First, the intersection point (IP) of the CoP trajectory for stride $s_k \in \tilde{S}$ is calculated: $CoPX_{ip}^{(s_k)}$, $CoPY_{ip}^{(s_k)}$. Then, the lateral and AP shift in the IP for a trial are given by:

$$\beta_L = \{\beta_L^{(s_k)} : s_k \in \tilde{S}\}, \; \beta_{AP} = \{\beta_{AP}^{(s_k)} : s_k \in \tilde{S}\}$$

Define $(CoPX_{ip}, CoPY_{ip}) = (\frac{\sum_{k=1}^{N_s} CoPX_{ip}^{(s_k)}}{N_s}, \frac{\sum_{k=1}^{N_s} CoPY_{ip}^{(s_k)}}{N_s})$ as the mean IP. The set of lateral and AP *squared deviation* from the mean IP for a trial are given by:

$$\alpha_L = \{\alpha_L^{(s_k)} = (CoPX_{ip}^{(s_k)} - CoPX_{ip})^2 : s_k \in \tilde{S}\}$$

$$\alpha_{AP} = \{\alpha_{AP}^{(s_k)} = (CoPY_{ip}^{(s_k)} - CoPY_{ip})^2 : s_k \in \tilde{S}\}$$

The lateral (η_L) and AP (η_{AP}) *asymmetry* can then be defined as the mean lateral and AP shift in the IPs, respectively. Similarly, the lateral (σ_L) and AP (σ_{AP}) *variability* are defined as the lateral and AP standard deviation (SD) in the IPs, respectively. We selected β_L and α_L as the two characteristic features of ML variability for our analysis.

Note that all features except the FPAs are always non-negative. Before deriving the stride-wise features, GCs with missing or invalid gait events were eliminated. Since several features, namely stride, swing times, stride length, width and angles will generate erroneous estimates for nonconsecutive strides, such values were dropped during data processing. Overall, 1654 (HOA: 905, PwMS: 749) and 1576 (HOA: 878, PwMS: 698) strides were retrieved from W and WT trials, respectively, across 35 subjects (HOA: 18, PwMS: 17).

TABLE I
SIZE-N NORMALIZATION FOR THE EXTRACTED GAIT FEATURES

Raw feature	L	T	F_z^e	C	SS	θ	W	P
Scaled feature	$\frac{L}{h}$	$\frac{T}{\sqrt{\frac{h}{g}}}$	$\frac{F_z^e}{wg}$	$\frac{C}{60\sqrt{\frac{g}{h}}}$	$\frac{SS}{\sqrt{gh}}$	θ	$\frac{W}{\frac{h}{60\sqrt{\frac{g}{h}}}}$	$\frac{P}{S_{size}}$

C. Data Normalization Techniques

The demographic differences between subjects may intrinsically influence the dynamics of gait variability and hence bias the MS gait differentiation efficiency. Thus, prior to classification, we normalized the subject's derived gait characteristics using the following two approaches:

1) Body Size-Based Dimensionless Normalization (Size-N): The extracted gait variables were normalized to non-dimensional forms by dividing via their corresponding dimension-matched body size-based scaling factors (proposed in [35]) in order to adjust for the inherent inter-subject physical differences. For instance, the acquired lengths, namely stride length and width were scaled by the subject's respective height. FPAs are dimensionless and thus require no scaling. Let w, h, S_{size} and g denote the body weight (in kg), height (m), shoe size (m) and acceleration of gravity (9.81m/s^2), respectively, then Table I summarizes scaled dimensionless quantities with regards to features obtained for both cohorts and trials where $L \in \{SL, SW\}$, $T \in \{SS_R, DS_R, DS_L, ST, S_tT, S_wT\}$, $F_z^e \;\; \forall \, e \in E$, $\theta \in \{\theta_L, \theta_R\}$ and $P \in \{\beta_L, \alpha_L\}$.

2) Multiple Regression-Based Normalization (Regress-N): Gait variables from both walking trials of the 35 subjects (in Section III-B) were normalized by regressing the baseline gait features of normative walking data from 30 additional healthy older adults on multiple demographic characteristics. These additional healthy older adults (age: 67.6 ± 10.34 years $[50 - 87$ years$]$, weight: 71.61 ± 14.52 kg $[52.97 - 103$ kg$]$, height: 1.68 ± 0.17m $[1.01 - 1.96$m$]$, male/female: $9/21$) were recruited from the local community. All controls walked for 200 s on the treadmill and yielded 21 gait features from a total of 3923 valid strides. A regression model was fitted to each gait feature with subject-wise averaged gait parameter values as a dependent variable and their corresponding demographics (weight, height, gender and age) as independent variables. All independent variables were assessed while fitting the regression since the variance inflation factor for each was lower than 5, hence ignoring the concern of multicollinearity. Further, the

Spearman's rank correlation coefficients among the independent variables presented no strong associations. For each gait feature, backward elimination was used to determine M statistically significant predictors ($p < 0.1$) and an optimal combination of predictors with the minimum corrected Akaike information criterion was selected out of 2^M possibilities. Subsequently, robust regression models minimizing the Tukey's biweight loss of the standard Gaussian residual errors were fit (see supplementary table S6 for the regression coefficients and the corresponding root mean squared errors). Gait features from both trials of the 35 study subjects (in Section III-B) were then normalized to dimensionless quantities with their predicted values obtained via their corresponding fit and subject demographics. Scaling relative to the regression predictors and coefficients computed from normative walking data of other healthy older adults aids in minimizing data spread among the gait features for the controls and association with individual demographic characteristics, and thus improve detection of MS vs. subject-related changes in gait.

D. Statistical Analysis

To examine cohort-related differences and the corresponding effect of normalization strategies on gait feature characteristics (i.e. mean, SD and range), a two tailed t-test and F-test was used to identify significant MS-related differences at $\alpha = 0.05$. The statistical assumptions of independence (since all subject observations were independent), normality (via the Shapiro Wilk test) and homoscedasticity (via Levene's test) were verified for the t-test. Mann-Whitney U-test and Welch's t-test were used, respectively, if normality or homoscedasticity, respectively failed. Similarly for the F-test, independence and normality were examined, and Levene's test was implemented if normality failed. Spearman's correlation (r) between the mean gait parameters and physical characteristics (weight, age, height and gender) of subjects in both trials were compared for raw (r_{raw}), body size (r_s), and regression (r_{reg}) normalized data to study the dependence of gait features (and thus the performance of ML models) on subject demographics. Further, among PwMS we explored the association and directionality of raw and normalized gait variables with disease severity using Spearman's correlations (r_{edss}) to motivate the applications of gait in learning MS progression with time.

E. Classification Models and Evaluation

MS prediction was studied across two classification designs, namely task and subject generalization (Fig. 1). In both task and subject generalization, binary supervised learning classifiers were trained to differentiate strides corresponding to HOA and PwMS. ML models were trained on 1654 strides across all 35 subjects in W trials and tested to categorize 1576 strides of the same subjects in WT trials for task generalization. Since our data set was limited to 35 subjects, we used a 7-fold cross-validation (CV) for subject generalization. In each scenario, all models were examined with both *size-N* and *regress-N* normalized features. Z-score normalization was applied to all features to eliminate the influence of variable feature ranges. For both classification architectures, the performance of nine notable supervised classifiers, i.e. decision tree (DT), random forest (RF), support vector machine with linear (LSVM) and radial basis function (RBF SVM) kernels, gradient boosting machine (GBM), adaptive boosting (AdaBoost), eXtreme gradient boosting (XGBoost), multilayer perceptron (MLP) and logistic regression (LR) were compared (see supplementary section S7 for details on these algorithms). Prediction efficiency for the task and subject generalization classifiers were weighed via the test set and mean CV precision, recall, accuracy, F_1 score and area under receiver operating characteristic (ROC) curve (AUC) metrics, respectively. Both setups were evaluated at stride and subject level categorizations, where majority voting was used to classify subjects into HOA vs. PwMS. Thus, a correctly classified subject's walk had more than 50% of strides accurately detected as of the appropriate cohort. Precisely, we annotate the stride and subject-level classification metrics with str (i.e. P_{str}, R_{str}, A_{str}, $\mathrm{F1}_{str}$, AUC_{str}) and sub (i.e. P_{sub}, R_{sub}, A_{sub}, $\mathrm{F1}_{sub}$, AUC_{sub}) in the subscript, respectively.

F. MS Progression Space

We attempt to describe the progression stage in PwMS by clustering their strides in distinct and multifaceted progression subgroups. Dimensionality reduction via rank-2 non-negative matrix factorization (NMF) was implemented on 21 *regress-N* features with 749 and 698 available strides of PwMS in trials W and WT, respectively to define a progression space for MS summarizing the influence of gait features in 2 dimensions (2D) across multiple stages. To impose non-negativity, all *regress-N* features were normalized between 0 and 1. Across both trials, NMF deconstructed the data into two matrices, namely progression vectors and the progression indicators. Progression vectors were used to construct the 2D *MS progression space (2D-MSPS)*. The 21 gait features were correlated to the two axes of the progression space using the magnitude of coefficients observed in the progression indicator vectors. Next, by applying unsupervised Gaussian mixture model (GMM) on the 2D-MSPS, we algorithmically parsed the progression space into three hidden subtypes within PwMS, representing the disease rate progressors. For each identified cluster, we study the number of strides and their share percentage in three severity subgroups (defined in Section III-B) based on the EDSS of MS subjects. Further, we look at the weights of the features to define a projection mapping for gait variables to the new 2D MSPS axes and thus find latent features describing the reduced progression space.

V. Experimental Results

Overall, PwMS reported longer and more dispersed stride, stance and double support times but a shortened single support on average in both the trials. Further, PwMS walked with a reduced stride length, cadence, self-controlled speed and a wider lateral distance between the two feet. PwMS reported higher median and spread in the BD extracted lateral shift (β_L) and squared deviation (α_L). In general, no individual or combination of features exhibit clear non-overlapping patterns characterizing MS. Any statistical model for MS prediction would thus be very high dimensional and prone to substantial scale and validation concerns. Therefore, ML-based investigation is an appropriate approach for the MS identification task.

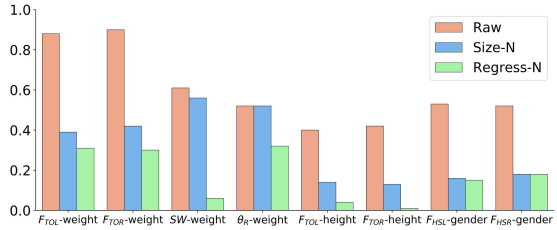

Fig. 3. Correlation with demographics. Absolute correlation of raw (red), *size-N* (blue) and *regress-N* (green) features with physical characteristics in trial W.

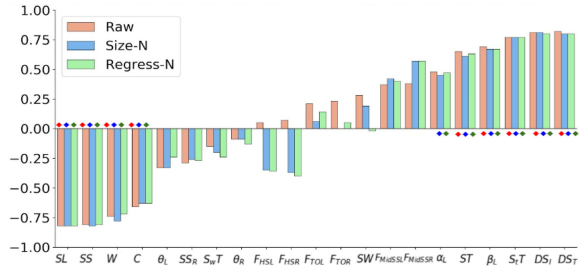

Fig. 4. EDSS Correlation. Bar plot illustrating the correlation of raw (red), *size-N* (blue) and *regress-N* (green) features with EDSS in trial W. Statistically significant correlations are marked with diamonds of respective colors.

A. Statistical Analysis

1) Statistical Significance: Subject-wise averaged raw and normalized features were compared between HOA and PwMS for significance of difference in means and variances. Considering trial W, statistically significant difference between means were observed in raw left FPA (6.4 times higher (6.4×) on average in HOA), lateral shift (1.7× in PwMS) and squared deviation (1.9× in PwMS). After the body size-based normalization, terminal double support (1.4× in PwMS), force on TOL (1.1× in PwMS), left FPA (6.4× in HOA), lateral shift (1.6× in PwMS) and squared deviation (1.9× in PwMS) demonstrated significance. When normalized using the regression technique, significant differences were noted in terminal double support (1.4× in PwMS), lateral shift (1.6× in PwMS) and squared deviation (1.8× in PwMS). With respect to trial WT, only raw terminal double support (1.5× in PwMS) and lateral shift (1.6× in PwMS) were significant and using the *size-N* data, terminal double support (1.5× in PwMS), lateral shift (1.6× in PwMS) and squared deviation (2.7× in PwMS) exhibited statistical significance. Similar to *size-N*, *regress-N* terminal double support (1.5× in PwMS), lateral shift (1.6× in PwMS) and squared deviation (2.6× in PwMS) showed significance in trial WT. 8 raw, 10 *size-N* and 12 *regress-N* features in trial W and 11 raw, 14 *size-N* and 14 *regress-N* features in trial WT indicated significant differences between variances (see supplementary table S8 for the list). In essence, both the normalization increased the number of parameters that exhibit significant difference between means and variances of the two cohorts.

2) Correlation With Physical Features: To explore the dependency of gait features on demographics, correlation (r) of physical properties with raw (r_{raw}), *size-N* (r_s) and *regress-N* (r_{reg}) parameters were compared. Across both trials, the range of correlations with raw data (W: $-0.41 \leq r^W_{raw} \leq 0.91$, WT: $-0.46 \leq r^{WT}_{raw} \leq 0.89$) lowered with *size-N* ($-0.46 \leq r^W_s \leq 0.56$, $-0.49 \leq r^{WT}_s \leq 0.53$) and further declined with *regress-N* features ($-0.41 \leq r^W_{reg} \leq 0.41$, $-0.44 \leq r^{WT}_{reg} \leq 0.51$). Fig. 3 plots some of these absolute correlations for trial W. For instance, *size-N* toe-off forces demonstrated significantly weaker correlations ($0.13 \leq |r_s| \leq 0.22$) with subject's height than their raw counterparts ($0.4 \leq |r_{raw}| \leq 0.43$). A similar trend was observed for the heel strike forces as well along with a further decrease for *regress-N* forces. High correlations between raw forces and subject's weight ($0.81 \leq |r_{raw}| \leq 0.91$) and gender ($0.42 \leq |r_{raw}| \leq 0.62$) weakened considerably with *size-N* to $0.03 \leq |r_s| \leq 0.46$ and $0.12 \leq |r_s| \leq 0.19$, respectively and with *regress-N* forces to $0.01 \leq |r_{reg}| \leq 0.51$ and $0.01 \leq |r_{reg}| \leq 0.41$, respectively. Interestingly, interaction between single support and gender heightened from $0.1 \leq |r_{raw}| \leq 0.3$ to $0.24 \leq |r_s| \leq 0.41$ with size-based normalization. *Regress-N* weakened ($|r_{reg}| \leq 0.25$) most associations with very infrequently realizing moderate values ($0.25 < |r_{reg}| \leq 0.51$) over both trials. Specifically, prominent correlations between weight and stride width ($|r^W_{raw}| = 0.61$, $|r^{WT}_{raw}| = 0.62$), left FPA ($|r^{WT}_{raw}| = 0.46$) and right FPA ($|r^W_{raw}| = 0.24$, $|r^{WT}_{raw}| = 0.52$) distinctly lowered to $0.02 \leq |r_{reg}| \leq 0.32$. *Size-N* could not assist in diminishing these high associations between stride width, left/right FPA and weight. All high correlations ($|r| \geq 0.7$) reduced to moderate ($0.5 \leq |r| \leq 0.7$) or low ($|r| \leq 0.5$) values with normalization. Thus, normalization reduced the inherent subject specific differences associated with physical characteristics in the gait features, potentially enabling the ML models to focus on learning to differentiate only disease-specific characteristics present in the gait parameters and consecutively increase their test set generalizability.

3) Correlation With Disease Severity: To explore the association of gait parameters with severity among PwMS, correlation (r_{edss}) of EDSS with raw and normalized features was studied. Fig. 4 plots the correlations for trial W. The directionality of r_{edss} matched our instinct with speed, length and cadence inversely correlating; and stride, stance, double support times and lateral shift positively interacting with disability. With respect to all three data streams, EDSS showed the strongest negative correlations ($r_{edss} \leq -0.7$) with stride length and speed in both the trials, additionally with walk ratio in trial W and cadence in trial WT. The strongest positive correlations ($r_{edss} \geq 0.7$) were illustrated with double support and stance times in both trials and also with stride time in trial WT. Cadence in trial W and walk ratio in trial WT exhibited moderate negative associations ($-0.7 < r_{edss} \leq -0.5$). Moderate positive interactions ($0.5 \leq r_{edss} < 0.7$) were shown by lateral shift and only normalized forces at MidSSR in both trials as well as stride time in trial W and lateral deviation, force at MidSSL in trial WT. The computed correlations were statistically significant for nine raw and normalized parameters (SL, SS, C, W, ST, S_tT, DS_I, DS_T and α_L) in both trials and two additional variables (F_{MidSSL} and β_L) in trial WT. The correlation of forces at MidSSR demonstrated significance only after normalization. Significant correlations between gait characteristics and EDSS motivate the applications of gait in learning the progression space and clinical stages of MS.

TABLE II
TASK GENERALIZATION: STRIDE- AND SUBJECT-WISE TEST SET PERFORMANCE FOR TOP-5 ALGORITHMS

		Stride-based					Subject-based				
Algorithm	Data	Accuracy	Precision	Recall	F_1	AUC	Accuracy	Precision	Recall	F_1	AUC
RF	Size-N	0.743	0.730	0.666	0.697	0.841	0.829	0.923	0.706	0.80	0.935
	Regress-N	0.792	0.796	0.713	0.752	0.886	**0.943**	**1.0**	0.882	0.938	0.987
RBF SVM	Size-N	0.744	0.686	0.779	0.730	0.819	0.857	0.833	0.882	0.857	0.980
	Regress-N	0.785	0.720	**0.841**	0.776	0.868	**0.943**	0.941	**0.941**	**0.941**	0.997
GBM	Size-N	0.787	0.784	0.716	0.749	0.867	**0.943**	**1.0**	0.882	0.938	**1.0**
	Regress-N	**0.824**	**0.853**	0.729	**0.786**	**0.910**	**0.943**	**1.0**	0.882	0.938	**1.0**
XGBoost	Size-N	0.784	0.770	0.732	0.750	0.867	0.886	0.933	0.824	0.875	0.980
	Regress-N	0.815	0.827	0.735	0.778	0.901	0.914	**1.0**	0.824	0.903	**1.0**
MLP	Size-N	0.746	0.742	0.652	0.694	0.820	0.829	0.923	0.706	0.80	0.951
	Regress-N	0.795	0.854	0.648	0.737	0.878	0.886	**1.0**	0.765	0.867	0.974

B. Prediction Models

Nine classifiers were compared with *size-N* and *regress-N* data to categorize strides and subjects between HOA and MS cohorts for task (Section V-B1) and subject (Section V-B2) generalization.

1) Task Generalization: To examine the differences of single and dual-task walking on individual gait characteristics in older adults with and without MS, we used a linear mixed effects model. Overall, all individuals demonstrated a significant increase in stance time, initial and terminal double supports and forces at MidSSR and TOR, and a significant decrease in stride length and speed when going from W to WT trials. A significant two-way interaction between cohort and task indicates greater increases in stride, stance, swing and right single support times, stride length, speed and walk ratio for PwMS during WT trials compared to HOA during W trials. A significant decrease in stride width, cadence and forces at HSR, TOL, MidSSR and HSL was observed for PwMS in WT compared to HOA under W trials.

Table II summarizes the stride- and subject-wise evaluation metrics for top-5 task generalization classifiers on categorizing the test set strides of trial WT (see supplementary table S9 for hyperparameter exploration). Clearly, aggregated performance of all the subject's strides via majority voting improved upon the accuracy of individual stride-wise predictions, for instance from 74.3% to 82.9% and 79.2% to 94.3% on RF with *size-N* and *regress-N* data, respectively. The classification performances of all algorithms were higher across all metrics with the *regress-N* data except only for GBM with equal subject-wise metrics when using the *size-N* and *regress-N* data. LR, DT, linear SVM and AdaBoost are absent from Table II of top-5 classifiers. RF, RBF SVM and GBM achieved a subject classification accuracy (A_{sub}) of 94.3% with the *regress-N* data while A_{sub} for XGBoost and MLP were 91.4% and 88.6%, respectively with the regression normalized data. RF, RBF SVM, XGBoost and MLP resulted in an A_{sub} of less than 90% with the *size-N* data except GBM that matched the 94.3% accuracy of *regress-N*. The maximum stride classification AUC (AUC_{str}) was 0.91 followed by 0.90 using the *regress-N* data on GBM and XGBoost, respectively whereas the optimal AUC_{str} with the *size-N* data was 0.87 on GBM and XGBoost. RF, RBF SVM and MLP had an AUC_{str} of less that 0.85 when using the *size-N* data. Considering all evaluation metrics in Table II, GBM with *regress-N* data performed the best with an accuracy, F_1 and AUC of 82.4%, 0.79 and 0.91, respectively at stride-level and 94.3%, 0.94 and 1.0, respectively at subject-level classification, followed by RF and RBF SVM on *regress-N* with a matching subject-level accuracy. Boosting algorithms sequentially optimized the current DT by adapting to the errors on the data of prior weak learners as compared to RF training DTs in parallel on bootstrap samples, thus GBM significantly improved the performance of learners with low variance but high bias. Gradient boosters iteratively regress over negative gradients of any generic differentiable loss function to boost the weak learning DTs whereas AdaBoost reweighing the previously mistaken data points higher specifically optimizes an exponential loss. MLPs are efficient to form disconnected decision regions and learn any arbitrary complicated boundary, as suggested by the universal approximation theorem. The optimal task generalization algorithm was GBM trained on *regress-N* data with 150 boosting stages, depth of 7, learning rate of 0.15 and considered 5 features for checking the best split (see supplementary figure S10 for its confusion matrix). Only two PwMS were miss-classified as HOA.

2) Subject Generalization: Table III summarizes the mean and SD of 7-fold CV performance metrics for the top-5 subject generalization classifiers (see S9 for optimal hyperparameters). All algorithms except AdaBoost with regression normalization surpassed the diagnostic performance when using the standard size-based normalization. LR, linear/RBF SVM and XGBoost did not make it to top-5. The best mean A_{sub} was 80% (95% confidence interval (CI): [75, 85]) using the *regress-N* data with MLP while RF and MLP had the maximum A_{sub} of 57.1% with the *size-N* data. Overall in Table III, MLP with *regress-N* data performed the best with a mean accuracy, F_1 and AUC of 62.1%, 0.57 and 0.68, respectively at stride-level and 80%, 0.78 (95% CI: [0.72, 0.83]) and 0.86 (95% CI: [0.78, 0.93]), respectively at subject-level classification. Tree-based models handle highly correlated variables to avoid overfitting better than kernel SVM. Unlike traditional ML algorithms relying wholly on hand-crafted features, MLPs are capable of incrementally learning latent characteristics of the data and discover novel inherent feature hierarchies with increasing complexity of the design. Our optimal MLP architecture with 7 fully connected layers and ReLU non-linearity was trained for 200 epochs using the adaptive moment estimation (Adam) optimizer with an adaptive learning rate initially set to 0.001 and the cross entropy loss (see S10 for its confusion matrix). Four PwMS and three HOA got incorrectly classified. Thus, GBM achieved the best A_{sub} (94.3%) for task generalization, whereas MLP performed the best (80%) for subject generalization.

TABLE III
SUBJECT GENERALIZATION: STRIDE- AND SUBJECT-WISE MEAN CV PERFORMANCE FOR TOP-5 ALGORITHMS

		Stride-based					Subject-based				
Algorithm	Data	Accuracy	Precision	Recall	F_1	AUC	Accuracy	Precision	Recall	F_1	AUC
DT	Size-N	0.504±0.12	0.50±0.25	0.459±0.20	0.427±0.15	0.538±0.12	0.514±0.34	0.429±0.43	0.429±0.42	0.410±0.39	0.690±0.38
	Regress-N	0.541±0.08	0.526±0.22	0.512±0.19	0.467±0.11	0.597±0.11	0.60±0.24	0.476±0.38	0.50±0.38	0.462±0.34	0.679±0.27
RF	Size-N	0.533±0.16	0.547±0.28	0.418±0.24	0.408±0.19	0.635±0.23	0.571±0.25	0.548±0.34	0.548±0.33	0.514±0.29	0.69±0.27
	Regress-N	0.563±0.11	0.557±0.25	0.463±0.23	0.449±0.16	0.643±0.16	0.60±0.19	0.571±0.32	0.524±0.27	0.519±0.24	0.643±0.19
GBM	Size-N	0.538±0.18	0.557±0.29	0.453±0.25	0.434±0.20	0.617±0.22	0.486±0.28	0.333±0.36	0.429±0.42	0.371±0.38	0.726±0.29
	Regress-N	0.584±0.12	0.580±0.24	0.518±0.23	0.486±0.17	0.654±0.14	0.60±0.24	0.452±0.33	0.50±0.38	0.471±0.35	0.798±0.20
AdaBoost	Size-N	0.592±0.15	**0.595±0.23**	0.440±0.24	0.451±0.19	0.644±0.18	0.543±0.18	0.429±0.32	0.357±0.23	0.381±0.25	0.774±0.15
	Regress-N	0.586±0.10	0.562±0.28	0.432±0.19	0.459±0.17	0.598±0.19	0.60±0.21	0.524±0.38	0.452±0.33	0.467±0.32	0.631±0.30
MLP	Size-N	0.524±0.14	0.534±0.28	0.362±0.24	0.366±0.19	0.598±0.20	0.571±0.20	0.524±0.38	0.405±0.32	0.424±0.29	0.762±0.28
	Regress-N	**0.621±0.10**	0.579±0.22	**0.619±0.20**	**0.565±0.14**	**0.682±0.15**	**0.80±0.15**	**0.833±0.20**	**0.786±0.25**	**0.776±0.17**	**0.857±0.23**

TABLE IV
ABLATION STUDY: TASK AND SUBJECT GENERALIZATION MODELS

	Task generalization						Subject generalization					
Data	Best Algorithm	A_{sub}	P_{sub}	R_{sub}	F_{1sub}	AUC_{sub}	Best Algorithm	A_{sub}	P_{sub}	R_{sub}	F_{1sub}	AUC_{sub}
S	RF	0.74	0.75	0.71	0.73	0.83	GBM	0.63±0.17	0.64±0.44	0.32±0.22	0.41±0.28	0.54±0.19
T	XGBoost	0.80	0.86	0.71	0.77	0.91	MLP	0.66±0.21	0.71±0.22	0.33±0.45	0.45±0.29	0.61±0.26
K	GBM	0.83	0.92	0.71	0.80	0.92	MLP	0.69±0.21	0.69±0.33	0.63±0.37	0.61±0.31	0.77±0.25
ST	GBM	**0.94**	0.94	**0.94**	**0.94**	0.98	RBF SVM	0.63±0.20	0.57±0.49	0.26±0.23	0.36±0.31	0.56±0.16
S+K	RBF SVM	0.91	**1.0**	0.82	0.90	0.94	AdaBoost	0.71±0.18	0.81±0.35	0.51±0.35	0.58±0.30	0.71±0.25
T+K	MLP	0.91	**1.0**	0.82	0.90	0.98	AdaBoost	0.71±0.21	0.76±0.34	0.63±0.37	0.64±0.31	0.80±0.23
All	GBM	**0.94**	**1.0**	0.88	**0.94**	**1.0**	MLP	**0.80±0.15**	**0.833±0.20**	**0.786±0.25**	**0.776±0.17**	**0.857±0.23**

C. Post Hoc Analysis

Note that for further analysis, we adhered to only using *regress-N* data for it demonstrated superior performance across both task and subject generalization model designs.

1) Ablation Study: We compared the task and subject generalization performance on several subsets of *regress-N* features, namely 4 spatial (S), 7 temporal (T), 8 kinetic (K), 13 spatiotemporal (ST), 12 spatial-kinetic (S+K) and 15 temporal-kinetic (T+K) parameters, to that of using all 21 variables for MS prediction. All ML models were tuned from scratch on these data streams for comparison. Table IV illustrates the subject-wise metrics for the best performing algorithm on each subset across both the task and subject generalization schemes. Across both model designs, LR, DT and linear SVM were never the top performers. Overall, GBM and MLP followed by AdaBoost are the most prominent algorithms in Table IV for task and subject generalization, respectively. Task generalization revealed the best performance when using all 21 features with GBM (A_{sub}: 0.94, AUC_{sub}: 1.0) followed by spatiotemporal also with GBM (A_{sub}: 0.94, AUC_{sub}: 0.98) and temporal-kinetic parameters with MLP (A_{sub}: 0.91, AUC_{sub}: 0.98). For subject generalization, MLP with all features had the best mean results (A_{sub}: 0.80, AUC_{sub}: 0.86) followed by temporal-kinetic with AdaBoost (A_{sub}: 0.71, AUC_{sub}: 0.80) and spatial-kinetic also with AdaBoost (A_{sub}: 0.71, AUC_{sub}: 0.71). In both model designs, ML algorithms had a better performance using all features, thus these ablation results indeed support our decision to use all the extracted gait features for prediction.

2) Analysis of Feature Importance: We first investigated the importance of features via *conditional entropy* (CE). The CE of labels Y, taking binary values, with respect to the discretized feature X, taking values in a finite set

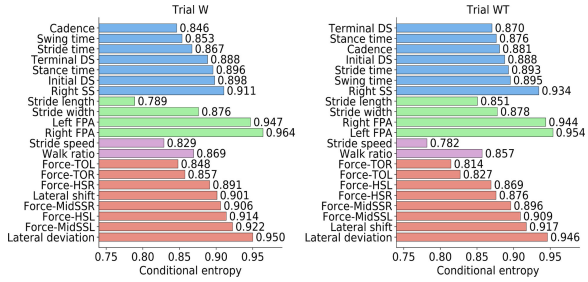

Fig. 5. The entropy present in the labels given *regress-N* gait features in trials W (left) and WT (right). Temporal, spatial, spatiotemporal and kinetic features are grouped in blue, green, plum and red colors, respectively.

X, was defined as: $\sum_{(x,y)\in X\times\{0,1\}} p_{X,Y}(x,y)\ln\frac{1}{p_{X,Y}(x,y)} - \sum_{x\in X} p_X(x)\ln\frac{1}{p_X(x)}$, where $p_{X,Y}$ is the joint probability mass function of (X,Y) and p_X is the probability mass function of X. Features with a low entropy reflect less randomness and hence are more predictive of labels. Fig. 5 depicts the CE of all features in trials W and WT. The most informative features with the least CE were (in order) $SL > SS > C > F_{TOL} > S_wT$ in trial W and $SS > F_{TOR} > F_{TOL} > SL > W$ in trial WT. Cadence followed by swing time in trial W and terminal double support followed by stance time in WT showed the most reduction in entropy among temporal features. Stride length followed by width from spatial, stride speed out of spatiotemporal and toe-off forces from kinetic features delivered the most predictive power in both trials. Overall, stride speed, length and forces at the toe-off were found to be the most valuable features across both trials. FPAs and lateral deviation with a high CE in both trials were least predictive of the labels. Given that our best ML algorithms, GBM and MLP for task and subject generalization,

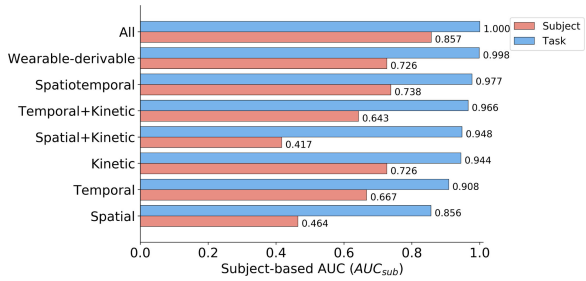

Fig. 6. Feature importance. AUC_{sub} for task and subject generalization models with different data domains are represented in blue and red, respectively.

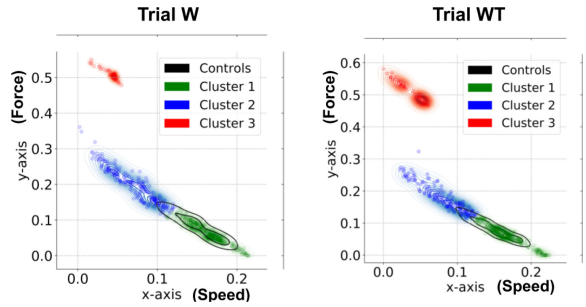

Fig. 7. Constructed 2D-MSPS. Left: Trial W, Right: Trial WT. Three clusters (shown in green, blue and red) are identified in strides of PwMS and distribution of HOA strides is depicted in black outlines.

respectively, used all 21 features, we also investigated feature importance by studying the decrease in performance of optimally tuned GBM and MLP models when only including features from specific subsets. Apart from subsets S, T, K, ST, S+K and T+K considered in Section V-C1, we defined another group as features obtainable from wearable sensors for this analysis. All defined gait features except the BD-based parameters could be derived from wearable foot switches or inertial sensors [36]. Fig. 6 depicts the AUC_{sub} for optimal task (GBM) and subject (MLP) generalization models with features from several data domains. For both models, using all features yielded the best AUC_{sub}, followed by wearable-derivable measures (0.998) and spatiotemporal (0.977) features for task generalization and by spatiotemporal (0.738) and wearable-derivable/kinetic (0.726) parameters for subject generalization. In both frameworks, no one set of features outperformed or matched the performance of using all features collectively. Especially for subject generalization, all features together are essential to diagnose the heterogeneity present in new subjects.

D. MS Progression Space

Promising correlations between gait features and EDSS (Section V-A3) motivated exploring gait-based characteristics to describe the MS progression space. To define hidden clinical subtypes within PwMS, unsupervised GMM was used to partition the NMF reduced 2D-MSPS. In both trials, three optimum number of underlying clusters for GMM were attained using the Bayesian information criterion (BIC). Fig. 7 depicts the three identified clusters in strides of PwMS with distribution in strides of controls superimposed for visualization in both trials. For each identified cluster, Table V summarizes the

TABLE V
COUNT AND RATIO OF STRIDES RELATIVE TO EDSS IN EACH CLUSTER. CLUSTERS 1, 2 AND 3 ARE ABBREVIATED AS C1, C2 AND C3, RESP

	Trial W			Trial WT		
EDSS	C1	C2	C3	C1	C2	C3
1.0-2.5 (mild)	131 (0.29)	17 (0.07)	0 (0.0)	149 (0.31)	12 (0.07)	0 (0.0)
3.0-4.5 (mild-to-moderate)	317 (0.69)	23 (0.09)	0 (0.0)	308 (0.65)	11 (0.06)	0 (0.0)
5.0-6.0 (moderate)	11 (0.02)	204 (0.84)	46 (1.0)	18 (0.04)	148 (0.87)	52 (1.0)

number of strides and their share percentage in three severity subgroups based on the EDSS of MS subjects. Cluster 1 (green) is dominated by strides of mild and mild-to-moderate severity patients. Cluster 2 (blue) is majority of moderate PwMS strides covering around 84% in trial W and 87% in WT of cluster observations and cluster 3 (red) has no mild or mild-to-moderate strides and contains only strides of moderate PwMS. The share of mild and mild-to-moderate strides is decreasing with an increase in the progression rate. Visually, distribution of control strides most overlaps with cluster 1 dominated by strides from mild and mild-to-moderate subgroups. Further, we looked at the weights of the 21 features to define a projection mapping for gait variables to the new 2D MSPS axes (see supplementary figure S11). For both trials, the horizontal axis was dominated by stride speed and its related components and vertical axis corresponded to force related features. Interestingly, gait speed and force measures were the top predictive power features too (as found in Section V-C2).

VI. DISCUSSION

This study examined MS and disability related changes in spatiotemporal and kinetic gait features after normalization; and evaluated the effectiveness of GML4MS to classify strides of PwMS from healthy controls, and generalize across different walking tasks and subjects after gait normalization. A few other works have explored ML to classify MS based on gait data. Gait features extracted from 3D ground reaction force data were adopted to discriminate healthy, cerebral palsy and MS subjects using two ML methods, namely nearest neighbours and MLP [22]. However, a very modest dataset with only four PwMS was employed for this study and thus limits the generalization of the classification results. Further, the study is limited in examining only force data and not exploring any tree-based ML algorithms. A recent study used smartphone and smartwatch sensors data and ML to distinguish among healthy controls, mildly ($PwMS_{mild}$) and moderately ($PwMS_{mod}$) disabled PwMS during a two-minute walk test [23]. Although this work investigates three well-known algorithms, namely, LR, SVM and RF to achieve the best accuracy of 82% differentiating $PwMS_{mod}$ from HOA and from $PwMS_{mild}$ and 66% identifying $PwMS_{mild}$ from HOA; the analysis on boosting algorithms, which have known to outperform RF in most applications, is missing. Moreover, our study utilizes up to 75 s of data for analysis, as compared to the longer data sample of two-minute walk in [23]. Another recent work analyzed a long short-term memory approach to classify fall risk in PwMS using accelerometers [37]. To the

best of our knowledge, this is the first study utilizing data driven ML for classification of individual strides of older PwMS using both spatiotemporal and kinetic features while walking. Our stride-based feature extraction approach derived multiple samples from a single subject, thus augmenting and introducing significant variations to our dataset to improve the generality of ML classifiers, which may allow for frequent and even real-time inferences.

The instrumented treadmill adopted for this study allowed for continuous gait monitoring of longer durations and distances within a compact footprint, relative to overground walking, and the capture of deviations from several successive strides [38]. While PwMS in this study were able to walk independently, the ceiling mounted harness, rails, and emergency stop provide essential tools for safety in PwMS with balance and fatigue concerns. Further, the integration of a built-in force plate supported kinetic data acquisition and allowed for online detection of gait events [31]. While walking on a treadmill can affect gait performance [39], these differences are generally within the normal variability of gait parameters and may be further diminished after an appropriate accommodation period to treadmill walking [40]. Our treadmill training before actual data collection and adaptive speed control helped subjects to more closely resemble natural walking.

Our work examined the benefits of regression normalized gait features on the accuracy of MS prediction using stride-based ML classification algorithms. Both the size- and regression-based normalization schemes increased the number of parameters demonstrating statistical significance between HOA and PwMS. The ability of *regress-N* normalization to reduce the association between gait features and personal demographics is crucial towards boosting the performance and generalizability of ML classifiers aimed at MS prediction. We have used statistical insights from admittedly a small number of test subjects. However, through the extraction of *regress-N* gait features, our approach mitigates some of the concerns related to small sample sizes since we are reducing the bias in the data by increasing independence (see Section V-A2). Compared to past studies on regression normalization in ML for other neurological disorders [21], [25] using the same controls in their classification set to extract regression coefficients, we used a normative dataset separate from our 35 study subjects to derive regression models for the gait features, hence prohibiting any divulgence of information from validation to training set.

Our proposed task generality framework demonstrates the feasibility of training on data collected in a lab-based walking task, and prediction on a walking while talking task, which paves the way for further inquiry into prediction using data collected in naturalistic and ecologically valid scenarios. We conclude that *regress-N* data with GBM and MLP were the optimal ML frameworks for task and subject generalization, respectively. An ablation study on the set of features supported using all the extracted gait features for better predictability in both model designs. From a clinical perspective, stride level classification allows for the use of a single stride, or brief duration walking trial, to serve as the basis for disease progression monitoring, which may be well suited for clinical settings with limited space and time. Further, as an effort towards the explainability of our ML-based study, we explored conditional entropy and decreases in performance of optimal GBM and MLP models. When only including a subset of features to examine the most relevant features driving the ML performance, we found that stride speed, length and forces at the toe-off were the most valuable features across both trials. Furthermore, we find that the use of wearable-derivable features is closely behind all features in terms of classification performance, which provides preliminary evidence of the feasibility of using wearable sensor data collected at home or local community in future telemedicine or rural health applications. Our study also examined how well normalized gait features could predict disability in PwMS. Significant correlations between gait characteristics and disability in PwMS (see Section V-A3) motivated the application of *regress-N* gait features in learning the progression space of MS (see Section V-D). Of particular significance, the two reduced dimensions arising after NMF were dominated by stride speed and force, which were also the most predictive features of MS-related changes.

The current work designs a domain knowledge-based MS screening model but the small cohort size recruited for this study limits making generalized interpretations for the heterogeneous MS community. Although, the features selected for predictive models in this study, namely, spatiotemporal characteristics (see [12]–[16]), FPAs [33], BD-based variables [12], [34] and forces [17], have been clinically shown and commonly adopted in the past to quantify gait impairments in PwMS, yet, by pre-selecting a specific set of domain knowledge-based features, we might be at a risk of introducing certain investigator bias in our ML models. Future work should focus on carefully characterizing the potentially missed information represented by the non-selected variables. ML explainability analysis in Section V-C serves as an initial estimate to demonstrate the influence of our feature selection on the model prediction performance. For an ideal understanding of dynamics from the inherently continuous gait data stream [41], we would need further exploration on non-linear dynamical features characterizing the human movement. Future research should examine associations of gait parameters with additional demographic and clinical factors to design improved normalization techniques. Further evaluation of GML4MS on a separate MS dataset with additional concurrent tasks, or while walking at home or in the community would be essential to establish robustness and improve sensitivity. Exploring hidden Markov and recurrent neural network predictive models by using tensors of independent strides will be vital to gauge the temporal component present in the continuous gait data. Future work is needed to identify prospective fall risk in MS subjects and assess the performance of our approach with remotely acquired gait data [23] and wearable sensors [37]. Further, observed correlations of gait parameters with disability may help identify older PwMS advancing into sudden worsening, which may provide improved personalized care, and merits future investigation.

VII. Conclusion

We present GML4MS, a novel ML pipeline for classification of PwMS using gait dynamics. The expression of MS over time and aging is heterogeneous, making the identification of sudden

changes in PwMS, particularly difficult. In this work, we extracted normalized spatiotemporal and kinetic gait features and demonstrated the benefits of *regress-N* to differentiate MS and disability related changes. Further, we evaluated the effectiveness of GML4MS to generalize across different walking tasks and subjects. With a larger data set, generalization of subjects in one test environment to new subjects in a different environment would need to be validated. The current study on prediction and progression space in MS may aid neurologists to understand advancing disease with aging and identify meaningful ML-based strategies for identifying PwMS. Given that we have more older adults with MS than younger adults, and the expected continual shift of the peak prevalence of MS into older age groups, the prediction of a tipping point for older PwMS advancing into sudden worsening may provide improved personalized care. Early detection of these inflection points in older PwMS may lead to concise and effective detection strategies and in turn benefit both patients as well as clinicians to curtail MS therapy expenses.

Acknowledgment

The authors would like to thank Gioella Chaparro and other members of the Mobility and Fall Prevention Research Lab for their assistance with data collection and the participants in this study for their contributions.

References

[1] I. Kister *et al.*, "Natural history of multiple sclerosis symptoms," *Int. J. MS Care*, vol. 15, no. 3, pp. 146–156, 2013.
[2] A. Henry *et al.*, "Anxiety and depression in patients with multiple sclerosis: The mediating effects of perceived social support," *Mult. Scler. Related Disord.*, vol. 27, pp. 46–51, 2019.
[3] A. Compston and A. Coles, "Multiple sclerosis," *Lancet*, vol. 372, no. 9648, pp. 1502–1517, 2008.
[4] C. L. Martin *et al.*, "Gait and balance impairment in early multiple sclerosis in the absence of clinical disability," *Mult. Scler. J.*, vol. 12, no. 5, pp. 620–628, 2006.
[5] K. Sharma *et al.*, "Epidemiology of Multiple Sclerosis in the United States," *Neurology*, 2018. [Online]. Available: http://n.neurology.org/content/90/15_Supplement/P1.140.abstract
[6] R. Marrie *et al.*, "The rising prevalence and changing age distribution of multiple sclerosis in Manitoba," *Neurol.*, vol. 74, no. 6, pp. 465–71, 2010.
[7] D. M. Hartung and D. Bourdette, "Addressing the rising prices of disease-modifying therapies for multiple sclerosis," *JAMA Neurol.*, vol. 76, no. 11, pp. 1285–1287, 2019.
[8] L. Scheinberg *et al.*, "Multiple sclerosis: Earning a living," *N. Y. State J. Med.*, vol. 80, pp. 1395–400, 1980.
[9] J. Noseworthy *et al.*, "Multiple sclerosis," *New England J. Med.*, vol. 343, no. 13, pp. 938–952, 2000.
[10] B. M. Sandroff *et al.*, "Relationships among physical inactivity, deconditioning, and walking impairment in persons with multiple sclerosis," *J. Neurol. Phys. Ther.*, vol. 39, no. 2, pp. 103–110, 2015.
[11] L. J. Balcer, "Clinical outcome measures for research in multiple sclerosis," *J. Neuro-Ophthalmol.*, vol. 21, no. 4, pp. 296–301, 2001.
[12] A. Kalron *et al.*, "Quantifying gait impairment using an instrumented treadmill in people with multiple sclerosis," *ISRN Neurol.*, pp. 1–6, 2013, Art. no. 867575.
[13] L. Comber *et al.*, "Gait deficits in people with multiple sclerosis: A systematic review and meta-analysis," *Gait Posture*, vol. 51, pp. 25–35, 2017.
[14] G. Severini *et al.*, "Evaluation of clinical gait analysis parameters in patients affected by multiple sclerosis: Analysis of kinematics," *Clin. Biomech.*, vol. 45, pp. 1–8, 2017.
[15] I. Dujmovic *et al.*, "Gait pattern in patients with different multiple sclerosis phenotypes," *Multiple Scler. Related Disord.*, vol. 13, pp. 13–20, 2017.
[16] M. Psarakis *et al.*, "Wearable technology reveals gait compensations, unstable walking patterns and fatigue in people with multiple sclerosis," *Physiol. Meas.*, vol. 39, no. 7, Jul. 2018, Art. no. 075004.
[17] R. Kaur *et al.*, "Exploring characteristic features in gait patterns for predicting multiple sclerosis," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Jul. 2019, pp. 4217–4220.
[18] K. J. Kelleher *et al.*, "The characterisation of gait patterns of people with multiple sclerosis," *Disabil. Rehabil.*, vol. 32, no. 15, pp. 1242–1250, 2010.
[19] J. P. Kaipust *et al.*, "Gait variability measures reveal differences between multiple sclerosis patients and healthy controls, "*Motor Control*, vol. 16, no. 2, pp. 229–244, 2012.
[20] N. M. Tahir and H. H. Manap, "Parkinson disease gait classification based on machine learning approach," *J. Appl. Sci.*, vol. 12, no. 2, pp. 180–185, 2012.
[21] F. Wahid *et al.*, "Classification of Parkinson's disease gait using spatial-temporal gait features," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 6, pp. 1794–1802, Jun. 2015.
[22] M. Alaqtash *et al.*, "Automatic classification of pathological gait patterns using ground reaction forces and machine learning algorithms," in *Proc. EMBC. IEEE*, 2011, pp. 453–457.
[23] A. P. Creagh *et al.*, "Smartphone- and smartwatch-based remote characterisation of ambulation in multiple sclerosis during the two-minute walk test," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 3, pp. 838–849, Mar. 2021.
[24] A. Kyvelidou *et al.*, "Aging and partial body weight support affects gait variability," *J. Neuroeng. Rehabil.*, vol. 5, no. 1, pp. 22–23, 2008.
[25] J. Kamruzzaman and R. K. Begg, "Support vector machines and other pattern recognition approaches to the diagnosis of cerebral palsy gait," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2479–2490, Nov. 2006.
[26] J. L. Preiningerova *et al.*, "Spatial and temporal characteristics of gait as outcome measures in multiple sclerosis (EDSS 0 to 6.5)," *J. Neuroeng. Rehabil.*, vol. 12, no. 1, pp. 14–21, 2015.
[27] A. Kalron, "Gait variability across the disability spectrum in people with multiple sclerosis," *J. Neurol. Sci.*, vol. 361, pp. 1–6, 2016.
[28] I. Hillel *et al.*, "Is every-day walking in older adults more analogous to dual-task walking or to usual walking? Elucidating the gaps between gait performance in the lab and during 24/7 monitoring," *Eur. Rev. Aging Phys. Activity*, vol. 16, no. 1, pp. 6–18, 2019.
[29] J. Verghese *et al.*, "Validity of divided attention tasks in predicting falls in older individuals: A preliminary study," *J. Amer. Geriatrics Soc.*, vol. 50, no. 9, pp. 1572–1576, 2002.
[30] Cuefors 2, ForceLink, Culemborg, The Netherlands, [Online]. Available: https://www.motekmedical.com/solution/c-mill/
[31] M. Roerdink *et al.*, "Online gait event detection using a large force platform embedded in a treadmill," *J. Biomech.*, vol. 41, no. 12, pp. 2628–2632, 2008.
[32] J. F. Kurtzke, "Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS)," *Neurol.*, vol. 33, no. 11, pp. 1444–1444, 1983.
[33] A. Karatsidis *et al.*, "Validation of wearable visual feedback for retraining foot progression angle using inertial sensors and an augmented reality headset," *J. Neuroeng. Rehabil.*, vol. 15, no. 1, pp. 78–90, 2018.
[34] A. Kalron and L. Frid, "The butterfly diagram: A gait marker for neurological and cerebellar impairment in people with multiple sclerosis," *J. Neurol. Sci.*, vol. 358, no. 1-2, pp. 92–100, 2015.
[35] A. L. Hof, "Scaling gait data to body size," *Gait Posture*, vol. 3, no. 4, pp. 222–223, 1996.
[36] S. Chen *et al.*, "Toward pervasive gait analysis with wearable sensors: A systematic review," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 6, pp. 1521–1537, Sep. 2016.
[37] B. M. Meyer *et al.*, "Wearables and deep learning classify fall risk from gait in multiple sclerosis," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 5, pp. 1824–1831, May 2021.
[38] S. Papegaaij and F. Steenbrink, "Clinical gait analysis: Treadmill-based vs overground," *Motek Inc.*: Amsterdam, The Netherlands, May 2017. [Online]. Available: https://knowledge.motekmedical.com/wp-content/uploads/2019/04/Motek-White-Paper-Clinical-Gait-Analysis.pdf
[39] H. Stolze *et al.*, "Gait analysis during treadmill and overground locomotion in children and adults," *Electroencephalogr. Clin. Neurophysiol./Electromyogr. Motor Control*, vol. 105, no. 6, pp. 490–497, 1997.
[40] P. O. Riley *et al.*, "A kinematic and kinetic comparison of overground and treadmill walking in healthy subjects," *Gait Posture*, vol. 26, no. 1, pp. 17–24, 2007.
[41] P. Federolf *et al.*, "A holistic approach to study the temporal variability in gait," *J. Biomech.*, vol. 45, no. 7, pp. 1127–1132, 2012.