



Human-centric algebraic machine learning

AML based world models - Revised

Deliverable D5.5

**WP5 - Learning World Models, Ethics
and Culture with AML**



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952091.

Consortium



ALGEBRAIC AI



Chantalimaud
Foundation



Project

Title: ALMA: Human Centric Algebraic Machine Learning
 Acronym: ALMA
 Coordinator: Proyectos y Sistemas de Mantenimiento SL (eProsima)
 Grant agreement: 952091
 Call: H2020-EIC-FETPROACT-2019
 Program: Horizon 2020
 Start: 01 September 2020
 Duration: 48 months
 Website: <https://alma-ai.eu>
 E-mail: alma@eprosima.com
 Consortium: **Proyectos y Sistemas de Mantenimiento SL (eProsima) (EPROS)**, Spain
Champalimaud Foundation (CF), Portugal
ALGEBRAIC AI (ALG), Spain
German Research Center for Artificial Intelligence (DFKI), Germany
Inria (Inria), France
TU Kaiserslautern (TUK), Germany
FIWARE Foundation e.V. (FF), Germany
Technical Research Centre of Finland (VTT), Finland
Carlos III University (UC3M), Spain

Deliverable

Number: **D5.5**
 Title: **AML based world models - Revised**
 Month: 18
 Work Package: WP5 - Learning World Models, Ethics and Culture with AML
 Work Package leader: DFKI
 Deliverable leader: Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI)
 Deliverable type: R: Document, report
 Dissemination level: Public (PU)
 Date of submission: 2022-02-28
 Version: v1.6
 Status: Finished

Version history

Version	Date	Responsible	Author/Reviewer	Comments
v1.0	2021-11-28	VTT	Stephen Fox	Partial Draft
v1.1	2021-12-18	VTT	Stephen Fox, Vitor Fortes Rey, Janin Koch	Partial Revision
v1.2	2021-12-20	VTT	Stephen Fox, Vitor Fortes Rey, Janin Koch	Expanded Draft
v1.3	2022-02-16	VTT	Stephen Fox, Vitor Fortes Rey	Complete Draft
v1.4	2022-02-23	DFKI	Vitor Fortes Rey	Embedding and results Revision
v1.5	2022-02-23	DFKI, VTT, Inria	Vitor Fortes Rey, Janin Koch	
v1.6	2022-02-28	DFKI, VTT, Inria	Vitor Fortes Rey, Janin Koch	Finished

Executive summary

In this document, Deliverable D5.5 - AML based world models - Revised, we report findings from our continued exploratory research into the potential of Algebraic Machine Learning (AML) to support ethical world models.

In the previous, Deliverable D5.4 - AML based world models - Initial, we made a start in addressing interfaces between individuals' embodied models of themselves in the world and organizations' documented models of themselves in the world. In particular, interfaces between individuals who have minor skin maladies and organizations that provide care to individuals who have minor skin maladies. The interface we investigated was AML-supported analyses of skin malady scans.

In this Deliverable, D5.5 - AML based world models - Revised, we continue with exploring interfaces between individuals' embodied models and organizations' documented models. Here, the interface between individuals' embodied models and organizations' documented models is AML-supported analyses of gait features captured with cameras and/or sensors.

In particular, our research reported in this document, D5.5, explored the potential of AML improve upon the explainability of AI-supported analyses of gait features. In particular, interpretability in terms of individuals' embodied models and organizations' documented models. This addresses the ethical requirement for people who may have disorders related gait to have explanation of what treatments are being recommended and are not being recommended by healthcare organizations. Also, it addresses the ethical requirement for healthcare providers to make the most effective use of scarce healthcare resources.

Our exploratory research indicates that AML may have some potential to improve classification based on gait features and the explainability of gait feature analyses because of its simple three level structure and conformance to human-based constraints specified in human-defined policies that can be written in AML-DL. Our research also indicates that AML at this point may have limitations regarding overfitting and complexity of models.

This deliverable, D5.5. AML based world models - Revised, will be followed by D5.8 AML based world models - Final. In addition, Work Package 5 will deliver D5.6 Embedding of ethics and cultural concerns into AML based world models - Initial, and D5.7 Embedding of ethics and cultural concerns into AML based world models - Final. In those future deliverables, we shall report our further research into the potential of Algebraic Machine Learning (AML) to improve interactive human-centric machine learning systems in complex settings.

Contents

Executive Summary	4
Contents	5
1 Introduction	7
2 Background	7
2.1 Gait	7
2.2 Gait analysis	8
2.3 Application of machine learning in gait analysis	9
2.4 Potential of AML to contribute to explainable gait analyses	10
3 Result from AML-analyses of gait features	11
3.1 Full List of Features	11
3.2 Discretisation AML Embedding	12
3.3 Training and Classification	13
3.4 Classification Results	13
3.5 Interpretability of AML models	15
4 Discussion	17
4.1 World models	17
4.2 Interfaces	17
4.3 Gait feature analyses as boundary objects	19
4.4 AML-enabled gait feature analyses as boundary objects	20
5 Conclusions	22
References	23

1 Introduction

In this deliverable, Deliverable D5.5 - AML based world models - Revised, we report findings from our continued exploratory research into the potential of Algebraic Machine Learning (AML) to support ethical world models. In the previous, Deliverable D5.4 - AML based world models - Initial, we made a start in addressing interfaces between individuals' embodied models of themselves in the world and organizations' documented models of themselves in the world. In this Deliverable, D5.5 - AML based world models - Revised, we continue with exploring interfaces between individuals' embodied models and organizations' documented models. Here, the interface between individuals' embodied models and organizations' documented models is AML-supported analyses of gait features captured with cameras and/or sensors. In particular, our research reported in this deliverable explored the potential of AML to improve the interpretability of AI-supported analyses of gait features. Specifically, interpretability in terms of individuals' embodied models and organizations' documented models. This addresses the ethical requirement for people who may have disorders related gait to have explanation of what treatments are being recommended and are not being recommended by healthcare organizations. Also, it addresses the ethical requirement for healthcare providers to make the most effective use of scarce healthcare resources. The remainder of the deliverable comprises four further sections: background, results, discussion, and conclusions.

2 Background

In this section, an overview is provided of the state-of-the art in gait feature analysis using machine learning (ML). First, the term, gait, is explained. Next, principal aspects of gait analysis are described. Then, application of machine learning to gait analysis is discussed. Subsequently, the potential of Algebraic Machine Learning (AML) to contribute to the explainability of gait analyses with machine learning is discussed.

2.1 Gait

The term, gait, refers to walking, running, and other means of natural locomotion combined with posture. A gait cycle consists of the activities that occur from the point of initial contact of one lower extremity, e.g. a foot, to the point at which the same extremity contacts the ground again. Gait cycle is a combined function of the lower extremity, pelvis, and spinal column. It is a single sequence of functions by one limb. It begins when reference foot contacts the ground and end with subsequent floor contact of the same foot. A single gait cycle is known as a stride [1]. Fundamental gait phases and expected interval of gait cycle are shown in Figure 1 below [2]. In particular, a single gait cycle consists primarily of two phases: a swing phase and a stance phase. Typically, the stance phase begins with the heel contact and ends with the toe of the same foot. That duration when foot remains in contact with the ground is known as stance phase and accounts for approximately 60 percent of the normal gait cycle.

The duration when the foot is off the ground is known as swing phase and accounts for 40 percent of the gait cycle. Swing phase begins with the toes of one foot and ends with the heel contact of that same foot. Each gait phase has a functional objective and a critical pattern of selective synergistic motion to accomplish its goal. The sequential combination of the phases also enables the limb to perform the following three basic tasks: weight acceptance, single-limb support, and limb advancement. The stance phase and the swing phase can be segmented further into eight segments: five stance and three swing. Specifically, the stance phase can be divided into five sub-phases as follows: initial contact, loading response, mid stance, terminal stance and pre-swing. The swing phase can be divided into the following three phases: initial swing, mid swing and terminal swing. Figure 1 illustrates the fundamentals of gait phases and expected interval of phases and sub-phases in the total gait cycle [2].

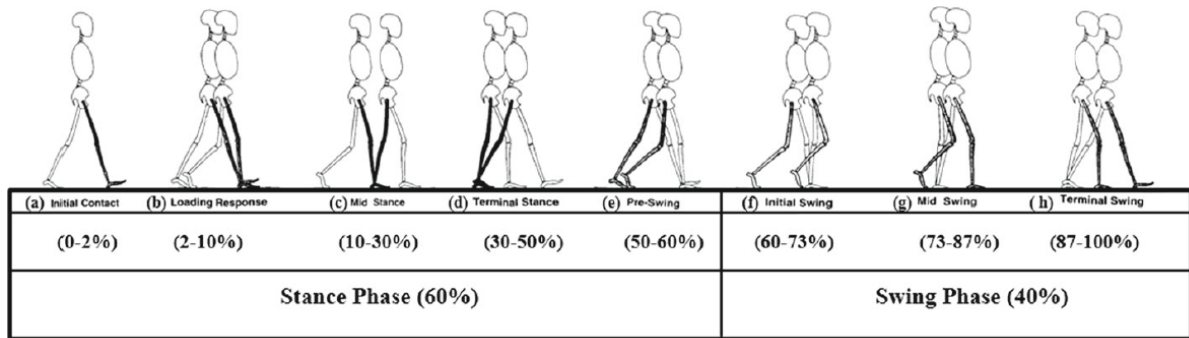


Figure 1: Fundamental gait phases and expected interval of gait cycle [2]

2.2 Gait analysis

Gait analysis includes the measurement of temporal, kinematics, kinetics and dynamic electromyography (EMG) based parameters from which conclusions about the subject (health, age, size, weight, speed, etc.) can be drawn [3]. Figure 2 shows a comprehensive parameter tree in gait analysis [1].

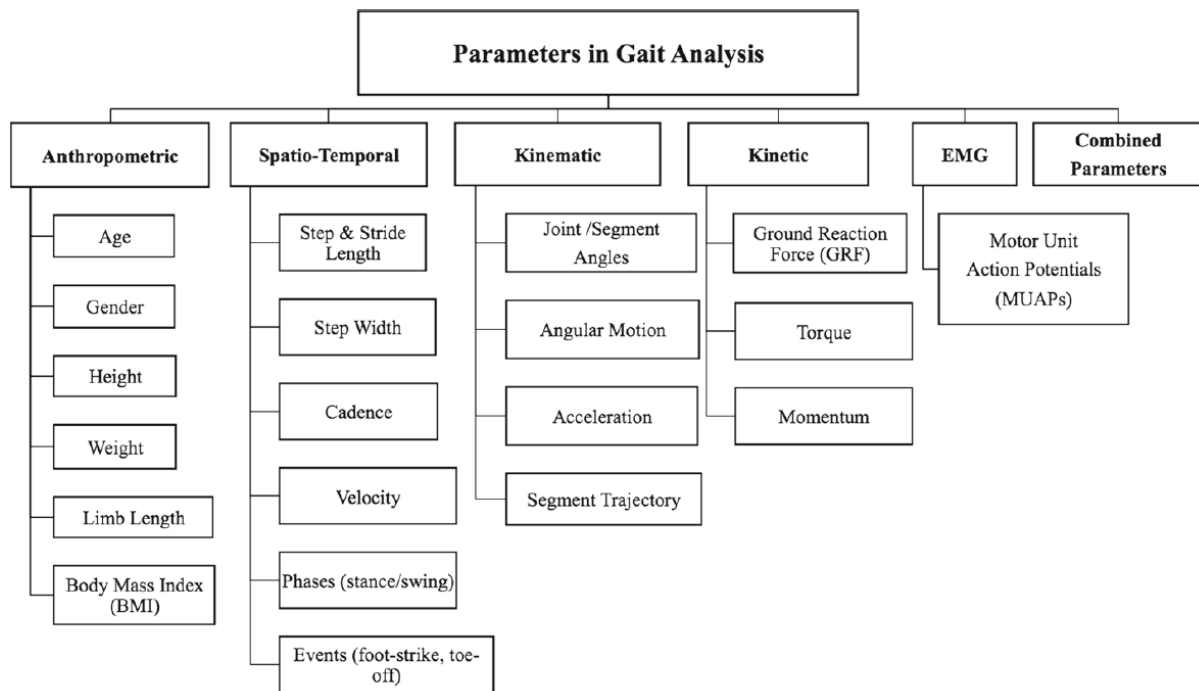


Figure 2: Gait analysis parameters [1]

As summarized in Figure 3, human gait analysis methods can be broadly classified into four types: vision based using a video camera (i.e. image processing based), sensor-based, other technologies, and combined approaches[1].

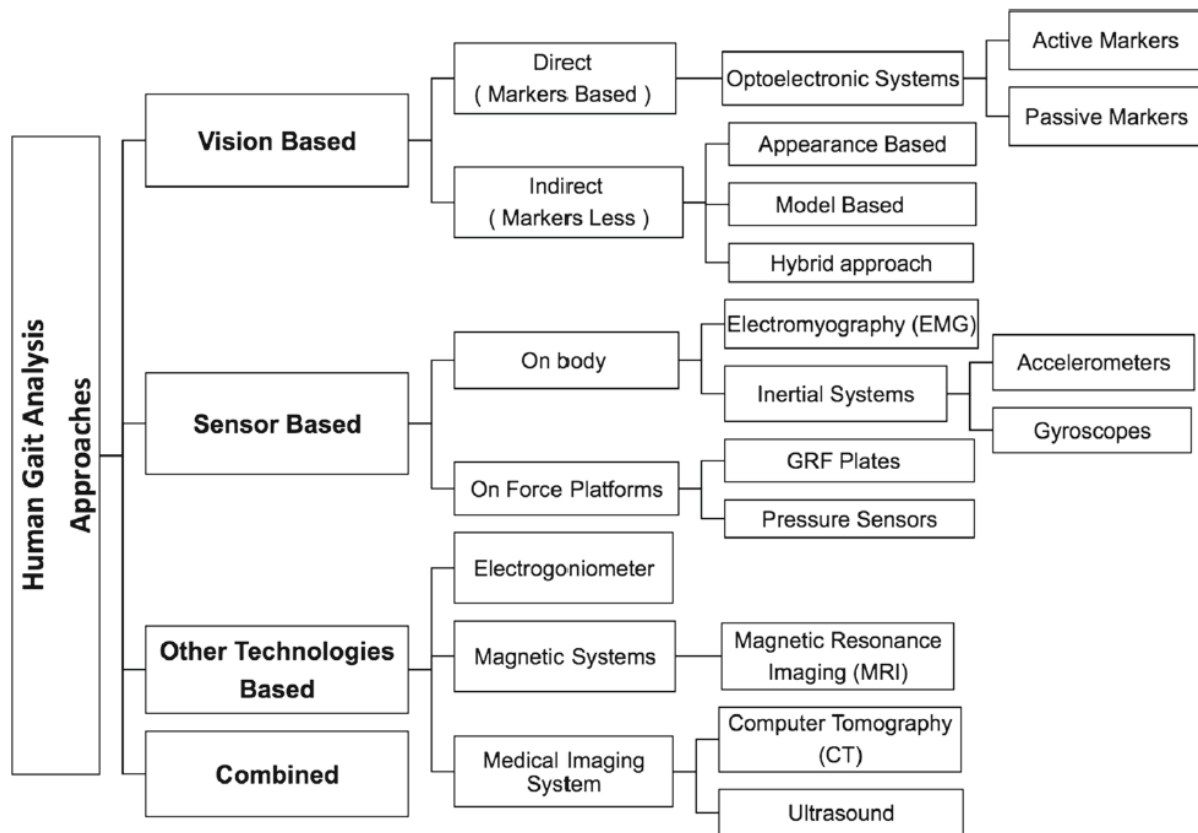


Figure 3: Gait analysis methods [1]

2.3 Application of machine learning in gait analysis

A range of machine learning techniques have been applied to gait analysis including the following: supervised (classification based), unsupervised (clustering based), reinforcement, rule-based, evolutionary (genetic algorithm, particle swarm optimization), probabilistic and hybrid approaches [1, 4, 5, 6]. Findings indicate that machine learning analyses have potential to differentiate between various gait-related disorders [7]. For example, one study focused on differentiating between physical impairments to gait, in particular functional impairments associated with the hip, the knee, the ankle, and the heel [8]. Another study proposed recognition of several health problems from gait analyses, including hemiplegia and Parkinson's disease [9]. A separate study applied machine learning successfully to assess condition within a functional disorder. In particular, within Complex Regional Pain Syndrome (CRPS): a condition that causes a long-term burning pain in a limb or part of a limb and can cause various degrees of physical functional performance deterioration [10]. A particularly interesting machine learning application in gait analysis is its use in detecting affective disorders, such as depression. It has been argued that gait analysis can provide a readily quantifiable objective approach to monitoring depression and related affective disorders [11]. This is important because regardless of what type of event may initiate disorders such as CRPS, i.e. medical conditions that impair normal functioning of bodily processes but that are not fully understood [12], they can be amplified and by perpetuated affective disorders [13]. For example, there can be strong links between anxiety, depression and fibromyalgia [14]. Apropos, machine learning has been applied to Kinect-recorded gait data in order to facilitate recognition of anxiety and depression [15]. However, it is important to note, that there have been few studies and little progress in improving the

explainability of results from gait analyses carried out with machine learning [16]. Accordingly, it is worthwhile to investigate the potential of AML to improve explainability.

2.4 Potential of AML to contribute to explainable gait analyses

AML-models are mathematical models comprising algebraic representations [17]. Algebraic Machine Learning (AML) is founded upon several core algorithms that can learn from data. Individual applications of AML to carry out specific analyses involve using AML description language (AML-DL) to define embeddings that include human-defined constraints. During the training, the core learning algorithms keep generating atoms until the model learns to differentiate between classes while fulfilling the human-defined AML-DL constraints. Thus, a valid model is obtained at the end of each training step. At this stage of the development of AML, the valid model may contain atoms that are not directly useful for classification, representing only general features of the problem. That is atoms which do not contribute to differentiating between classes while fulfilling human-defined constraints. Compared to machine learning models, such as some deep neural networks, which are recognized as less conducive to interpretability in the context of explainable AI (XAI) [18], AML models have three potential advantages.

First, AML-models have a fundamentally simple structure. They comprise only three layers: inputs – atoms – outputs. Relationships between inputs and outputs defined in atoms are binary. An atom can be linked to one or more inputs and is said to be present in an example if at least one of those inputs is present. AML-models expand as they learn patterns from training data. The number of atoms can increase from tens to thousands during training. Nonetheless, the simple three-layer structure persists.

Second, the core AML algorithms check that the human-defined constraints are maintained in the subsequent binary relationships between inputs and outputs that are learnt as new atoms are generated during training. To use an analogy, an initial AML-DL model description can be thought of metaphorically as a top-down development plan with predefined zone categories outlined for different types of development such as greenspace, housing, industrial, and retail. During training, the AML model will generate only new atoms from bottom-up sensory inputs that are consistent with the top-down development plan that is set-out in the initial AML-DL description. To continue with this analogy, the structure of the learned atoms would include bottom-up sensory inputs that are within predefined zone categories and exclude bottom-up sensory inputs that are not. Thus, the overall integrity of the AML-DL model is maintained however large the number of atoms grows to be.

Third, AML-DL rules can be traced back to human-defined policies. This is imperative because it must be clear from the outset and throughout what is included and what is excluded in an AML-DL defined model. For example, human-defined policies can provide the basis for human-defined work procedures. Then, work procedures can be distilled through a series of inter-related formats, such as natural language and summary diagrams, into first-order-logic rules that can define relationships in atoms between inputs and outputs in AML-DL. As new atoms generated during training conform to the AML-DL model's human pre-defined rules, AML-models are interpretable. That is, humans can interpret how the model uses inputs to generate outputs [19].

However, the expansion of AML-models from tens to thousands of atoms limits potential for direct human interpretation. This is because of the massive amount of time that would be required for people to go through every input – atom – output relationship. Accordingly, as described in the following section, we explore the potential of widely used decision tree methods to support human interpretation of AML models. The edges and nodes of tree structures provide natural visualization. Within the tree structure, data ends up in distinct groups that are often easier to understand than points on a multi-dimensional hyperplane as in linear regression. Furthermore, the explanations for short trees can be simple, because for each split the instance falls into either one or the other leaf. In other words, tree representations are binary. Thus, they are well-suited to AML because the relationships between inputs and outputs defined in AML atoms are binary. Yet, the interpretability of trees depends upon their depth. The more terminal

nodes and the deeper the tree, the more difficult it becomes to understand a tree [20]. Thus, we explore to what extent, if any, can application of classification and regression tree methods obtain interpretable summary information from the persistent simple structure of human-defined AML models.

3 Result from AML-analyses of gait features

In this work we have explored AML for Gait analysis using data from [21]. In that study, the gait characteristics of 121 healthy patients and 126 that were clinically diagnosed with depression was analysed. Gait was recorded using a Kinect (Version 2) while the subjects walked a 6 by 1 meter footpath, as depicted in Figure 4. Based on the joint as detected by the Kinect, [21] used statistical, frequency domain and spatiotemporal features to distinguish between the two groups of patients (clinically diagnosed and not) using different machine learning algorithms.

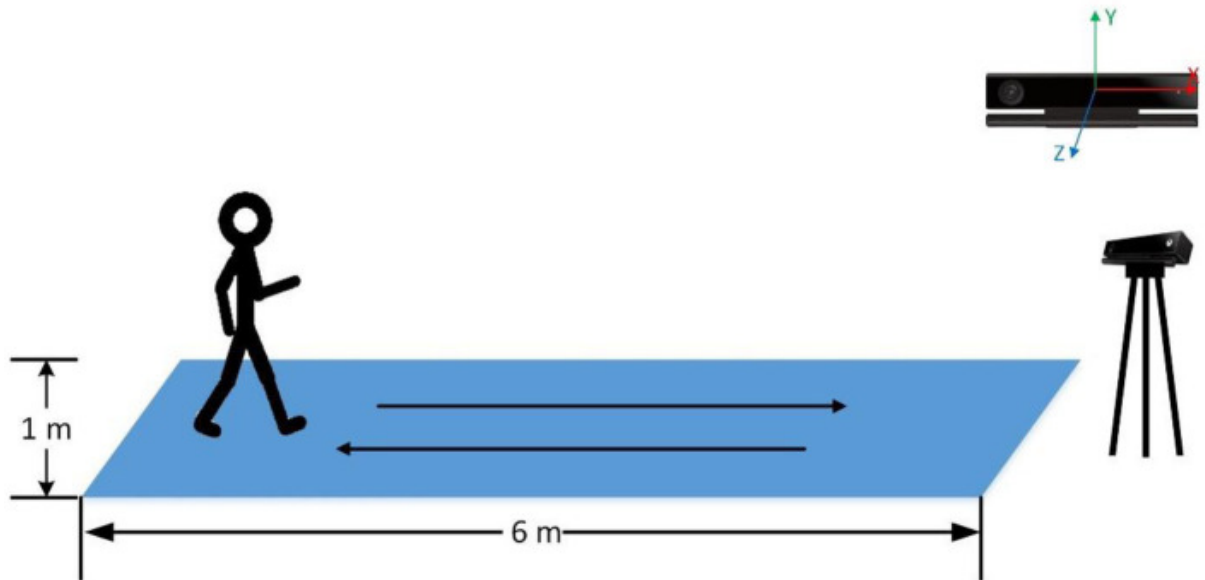


Figure 4: The schematic of the experiment environment [21]

While the best performance in this study is achieved with statistical and frequency domain features, those are only available pre-processed using principal component analysis, generating a feature set that is not interpretable. As our aim is explainable AI, we opted to use only the spatiotemporal ones instead, since they are not opaque and are related to previous research on gait patterns [21]. For example, they include walking speed, head posture and vertical head movement.

Thus, we will apply AML to this dataset and compare AML to other machine learning approaches.

3.1 Full List of Features

In [21], the spatiotemporal features computed are paraphrased in the following list:

- **Body swaying:** Body swaying is measured by the maximum difference in position of the left-shoulder and the right-shoulder on the X-axis during a gait cycle.
- **Left-arm/Right-arm swing:** It was defined as the maximum difference of the left/right wrist moving along the Z-axis during a gait cycle.

- Head posture: Quantifies the head posture during a gait cycle by averaging the angle between the vertical direction and the connection line between the neck and the clavicle in the plane which consisted of the Y-axis and the Z-axis.
- Left/Right stride length: It measured the maximum change in the horizontal direction of the left/right foot during a gait cycle.
- Walking speed: Participants' walking speed was measured according to their spine movement along the Z-axis.

In total, we have 10 features. For each subject, each feature is calculated using its mean during two gait cycles [21].

3.2 Discretisation AML Embedding

AML is based on Discrete Mathematics, so we need a strategy to discretize the data. In our case a simple approach was used, where each feature will be uniformly divided into 16 bins.

Regarding embedding, we used a simplified version of the one developed in D5.1, which expresses the notion of intensity of each value.

We define for each feature $30 = 2 \times (16 - 1)$ constants separated into two sets of 15 constants each, one set for “lower or equal” constants and another set for “greater than” constants. A total of $10 \times 30 = 300$ constants are used to describe the input and another 2 constants are used for the output, which represents the two groups of subjects.

For each feature f , the 30 constants associated to the sensor are $le_{0,f}, le_{1,f}, \dots, le_{14,f}$ and $g_{0,f}, g_{1,f}, \dots, g_{14,f}$. The formal knowledge we want to express is the notion of intensity, which we will represent as ordered sets

$$le_{i+1,f} \leq le_{i,f} \quad (1)$$

$$le_{i,f} \not\leq le_{i+1,f} \quad (2)$$

$$g_{i,f} \leq g_{i+1,f} \quad (3)$$

$$g_{i+1,f} \not\leq g_{i,f}, \quad (4)$$

forming one ascending chain and one descending chain for each feature. The goal is embedding the discretized feature values reading which corresponds to an intensity, i.e a numerical value and, therefore, represented as an ordered set in a discrete system.

Each training example can be represented as a vector $r(s)$ corresponding to the $s = 0, \dots, 10$ discretized feature readings for a subject and one output value k representing the group the subject belongs to. For each training example, a term T was created representing the problem input:

$$T = \odot_s \left\{ \begin{array}{ll} le_{0,s} & r(s) = 0 \\ le_{r(s),s} \odot g_{r(s)-1,s} & 0 < r(s) < 15 \\ g_{14,s} & r(s) = 15 \end{array} \right\} \quad (5)$$

and the training example was embedded with the positive relation

$$O_k \leq T \quad (6)$$

and 1 negative relation expressing that when some data belongs to one group he does not belong to the other,

$$O_{k'} \not\leq T. \quad (7)$$

The complete theory we have used to embed the human activity problem then consists on how training data is labelled, **Eqs. 6** and **Eqs. 7**, and the formal knowledge relations **Eqs. 1**.

3.3 Training and Classification

Given the small size of the dataset, our strategy was to train two AML models, one to recognise the clinically depressed group M_k and another to recognise the healthy one $M_{k'}$.

To check whether a test example T belongs to a given class O_j , we check whether the relation $O_j \leq T$ holds. This relation holds in model M_j when all the atoms in constant O_j are a subset of the atoms of term T . We say an example T has m misses if there are m atoms missing from the set of atoms in O_j . Formally, this number of misses is the the discriminant of the relation $O_j \leq T$ in the output model M_j (see [17] for definitions),

$$m(T, j) = |\text{disc}_{M_j}(O_j, T)|$$

where O_j is the constant associated to output class j .

The assigned output class can then be calculated using

$$\text{output}(T) = \text{argmin}_j \left(\frac{m(T, j)}{|M_j|} \right), \quad (8)$$

where $|M_j|$ is the number of atoms in the model.

This is a heuristic that worked well in the previous deliverables such as D5.1. In the presence of a validation set better approaches exist based on comparing cumulative distributions of each class, but in this case we use this simple heuristic due to the size of the dataset and since this deliverable is focused on world models and interpretability and not on achieving the best possible classification performance.

3.4 Classification Results

Replicating the procedure in [21], we performed a 10-fold cross validation and compared AML to other machine learning algorithms. Results can be seen in Figure 5.

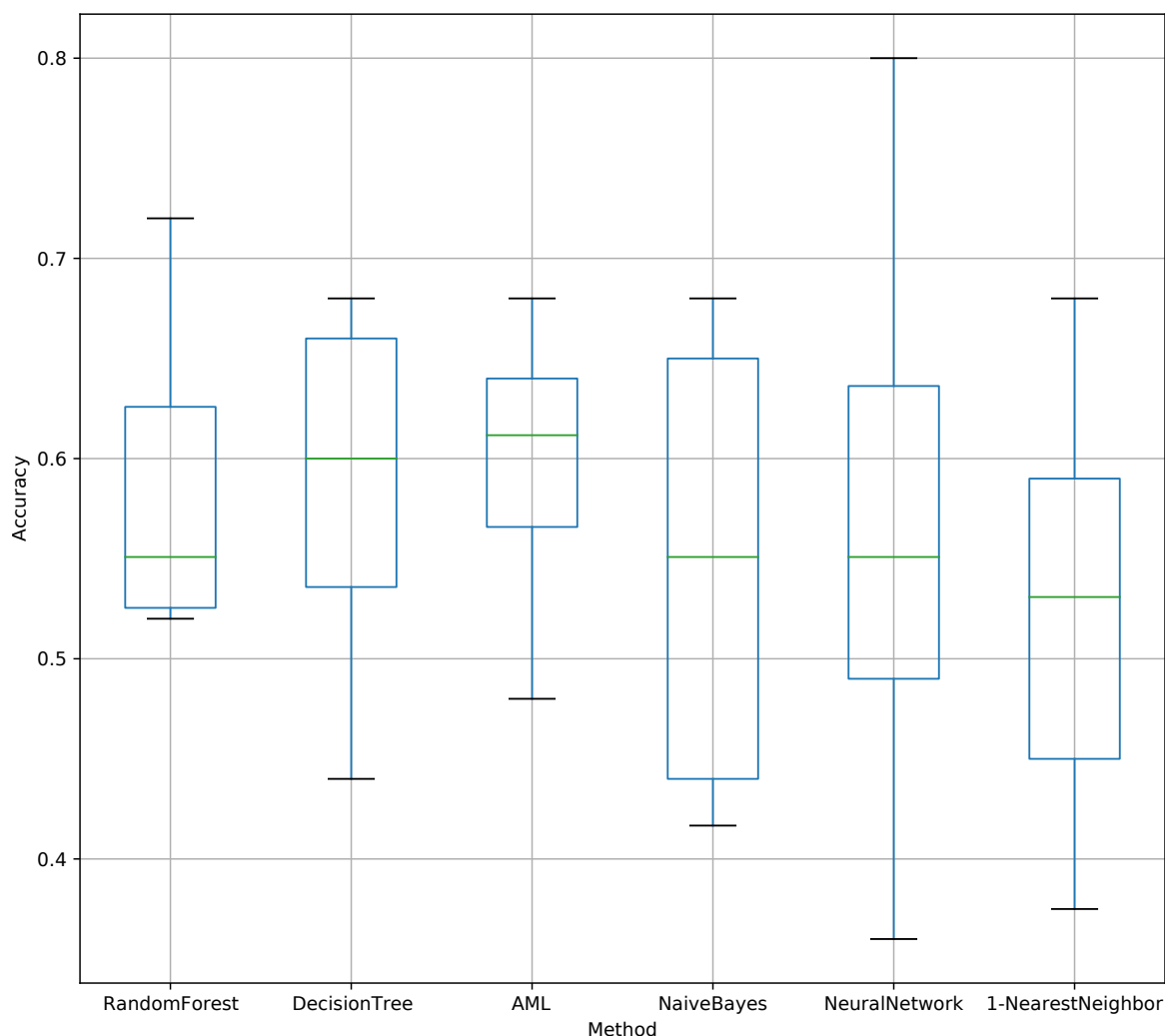


Figure 5: Results for the 10-fold cross validation using spatiotemporal features from [21] with different machine learning methods

As we can see, AML showed a better median classification accuracy than the other machine learning methods studied. Moreover, its result was also more consistent with a smaller standard deviation in its performance.

Notice that the number of atoms in the model can depend on the amount of time spent in training. In its current version AML models can grow its number of atoms as examples are revisited. As we are dealing with a small dataset, we stopped training once the training examples were all learned by the algebra. At this point the number of atoms is altogether around 6000, meaning we already have one order of magnitude more atoms than examples. As we can see in Figure 5, this did not lead to overfitting. Still, it is interesting to see what happens if we train our AML models further. Thus, we performed this experiment, training each models over several days. Results are depicted in Figure 6 where we show our original AML model as *AML* – 0 and test our model every half day of training time, with *AML* – 2 being the models after 1 day of training. It is clear that with our current methodology (heuristic of Section 3.3)

AML can degrade its performance, as seen in the difference between $AML - 0$ and $AML - 1$, but the overall performance seems to converge on a level above most of the other machine learning algorithms even after days of training.

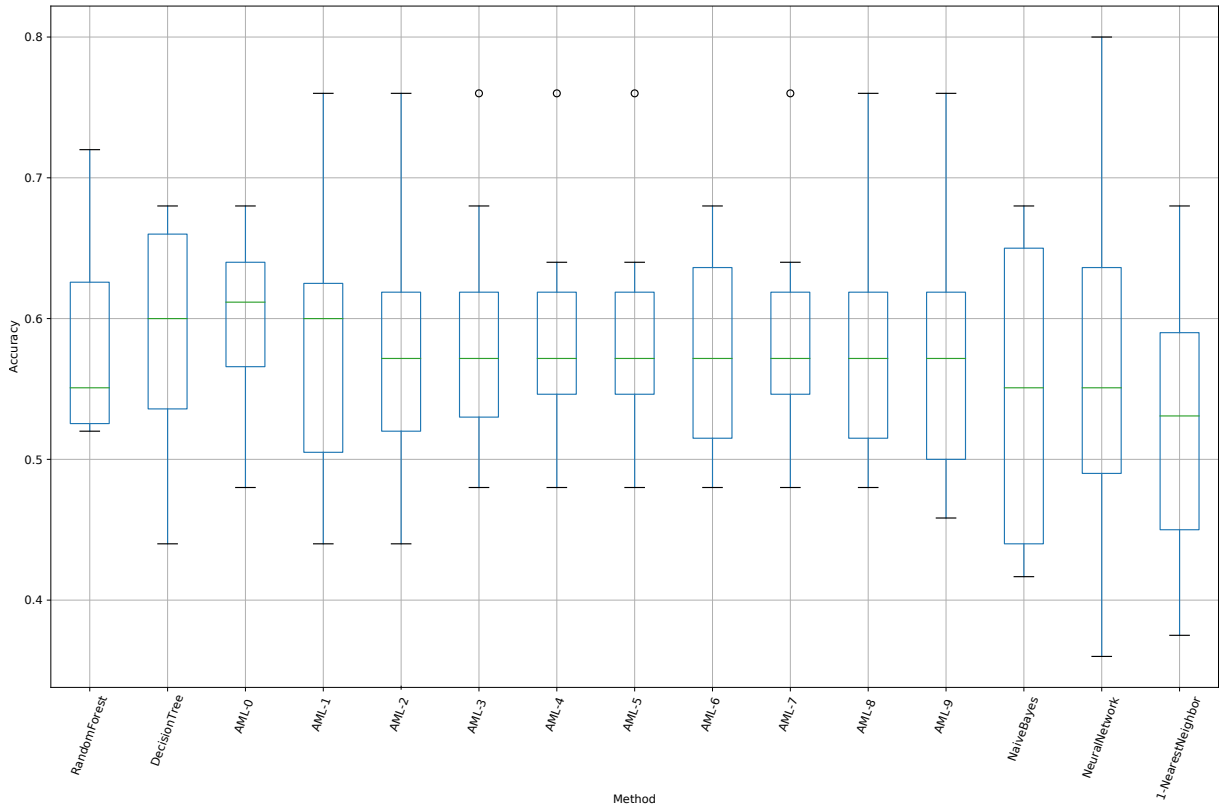


Figure 6: Results for the 10-fold cross validation using spatiotemporal features from [21] with different machine learning methods, now including longer training times for AML. We show our original AML model as $AML - 0$ and test our model every half day of training time, with $AML - 2$ being the models after 1 day of training.

3.5 Interpretability of AML models

As mentioned before, an AML model can have orders of magnitude more atoms than features. For example, with 10 features, the number of atoms can be in the thousands. Each atom can be easy to interpret individually. For example, one of the learned atoms is simply $headPosture > 1.18$, that is, this atom is present if this simple condition is fulfilled. On the other hand, some atoms include more conditions (being present if any of them is fulfilled) that can include many of the features.

Since looking manually at all the atoms for each decision is not feasible, one needs to present such information in a way that it can act as a boundary object.

The key idea is that not all atoms are relevant in a classification decision. For example, some atoms may always be present, or always be present when some other two atoms are present. Thus, in our first attempt we started by converting each example from the feature space to the atom space, that is, we convert every example to a vector x where

$$x_i = \begin{cases} 1, & \text{if atom } i \text{ is in } x \\ 0, & \text{otherwise} \end{cases}$$

This atom space representation can then be converted into a decision tree using the CART algorithm [22], limiting its height so that said tree is easier to interpret. An example can be seen in Figure 7.



Figure 7: Example of a tree as potential boundary object. Each node represents one atom with its listed set of presence criteria (one is enough for it to be included).

While this can reduce the amount of information one needs to analyse, and such representation could be interpreted by experts, it does not encode the full complexity of the AML model and is not intuitive for

non-experts.

4 Discussion

In this section, AML-enabled gait analyses are related to interpretability in terms of world models, interface states, and boundary objects. In particular, it is explained how different people's embodied models of themselves in the world can lead to them having different potential for gait disorders from the same initial circumstances. Also, it is explained that different organizations can have different perceptions of the same gait disorders. Accordingly, there can be different interpretations of exactly the same gait analysis report. Thus, improving the interpretability of machine learning-enabled gait analyses is insufficient in itself to lead to there being agreement in interpretations. This discussion section highlights that the current interpretability of AML-enabled gait analyses, as shown in Figure 7, is not yet adequate. This is because the tree, its branches, and its leaves are not readily accessible to non-experts. Hence, they are open to different interpretations by different people in accordance with their existing world models, which can be in opposition to each other. This is a very important issue in functional disorders, because different interpretations can lead to high treatment costs with poor treatment outcomes.

4.1 World models

The world models of individuals and organizations are models of self in the world. In human self-models, the mind is an embodied system in the world rather than just a neural network in the head. For individual people, embodied models of self in the world include implicit characteristics that provide internal context for processing inputs from the world and for taking actions in the world [23, 24, 25]. Implicit characteristics can include personality type, hardiness levels, fascia system, and body memory [26, 27, 28, 29]. These characteristics are difficult to categorize individually. For example, boundaries between personality types are difficult to define, and interrelationships between different characteristics can be uncertain [30, 31, 32]. By contrast, organizations make their internal characteristics explicit in order to demonstrate compliance with relevant laws, regulations, standards, etc. For example, organizations make their internal structure explicit in organograms and their internal beliefs clear in policy statements within the documentation of their quality management systems [33, 34, 35].

Embodied self-models provide individual people with prior expectations of what they will experience. Documented organization-models provide organizations with prior expectations of what they will experience. For example, people who expect to experience pain are more likely to experience pain [36]. What organizations expect to experience can be recorded in documents, such as plans for the coming years [37]. Individual people who pay more attention to their embodied models of themselves in the world than to sensory inputs from the world can engage in motivated cognition [38] and wishful seeing [39]. Similarly, organizations' lock-ins to their internal models can lead to persisting with path dependent actions even when there is sensory evidence that they are no longer effective [40, 41]. In summary, motivated cognition and wishful seeing, lock-ins and path dependencies can lead to information being interpreted to serve explanations that support preconceptions [42] and confirm biases [43]. For example, the tree shown in Figure 7 could be so interpreted.

4.2 Interfaces

Relationships between changing internal models and a largely unchanging external world is illustrated in Figure 8. Here, the world is largely unchanging in relation to the daily life of an individual or an organization. The shaded area highlights the extent of information gaps at interfaces between internal model and external world. Initially, at t_1 , there are minimal information gaps (green interface). However, the internal model changes and information gaps at interfaces increase until they are not sustainable (red interface) at t_3 .

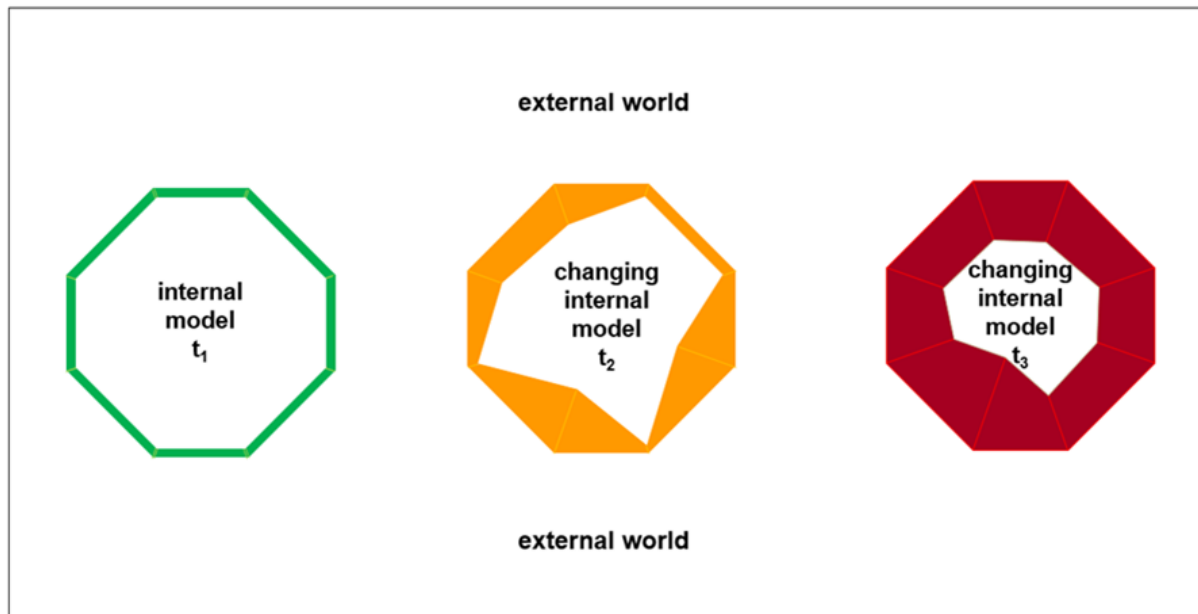


Figure 8: Interfaces of changing internal models

Consider, for example, changing internal models of human production operatives in a largely unchanging work place and how these can affect gait. A human production operative who works with mobile robotics while wearing hardware, such as exoskeletons and headsets, may develop a consistent top-down prior expectation of ankle pain from what was initially a surprise ankle pain early in a working day. This could have arisen from wearing an exoskeleton having introduced unpredictable loads to the production operative's musculoskeletal system while restricting three-dimensional rotational movements that are typically involved in lifting. Toward the end of the day, there may be very strong bottom-up sensory signals because the operative has to do some heavy lifting. If the operative pays attention to the ankle, then it is likely that the operative will be conscious of ankle pain, and gait may be affected. However, as the end of the working day nears, the operative's attention to ankle pain can be modulated by attention being focused on leaving work on time to keep a social appointment. As a consequence, the operative may not be conscious of ankle pain at the final stages of the day despite there being several hours of consistent prior evidence to support ankle pain expectation and despite strong physical causation for increased ankle pain during heavy lifting [44, 45].

Whether conscious or not conscious of ankle pain at the end of the working day, the operative may subsequently undergo some minor ankle fascia manipulation therapy and begin one ankle rehabilitation exercise to widen the range of ankle movement. After a few days, the ankle may be pain-free, the operative may soon forget about having had ankle pain, and the operative's gait may return to its previous pattern. However, the wearing of exoskeletons could have negative effects on fascia system functioning [46], such as negative unintended changes to interoception, nociception, and proprioception. Apropos, the next day, the operative's anxiety about ankle injury could lead the operative to becoming hypervigilant to any bottom-up sensory signal related to the ankle. This can involve two attentional processes of hypervigilance: detection of threatening stimuli and difficulty in disengaging attention from threatening stimuli. In particular, operatives with high pain-related anxiety are more likely to orient their attention toward a pain-related threat and have difficulty in disengaging from the threat [47]. Treatment may not end hypervigilance. Rather, treatment can be followed by fear of re-injury and prevent the return to work[48]. Accordingly, the operative's gait can change. At worst, there can be descent into

chronic fear of pain from movement: i.e., kinesiophobia, and chronic pain conditions with widespread complex symptoms such as fibromyalgia [49] or other functional disorder [13]. Here, it is important to note chronic body-wide pain is prevalent amongst many populations [50].

Thus, as summarized in Table 1, depending on the characteristics of individuals' embodied self-models, emergence of prior expectation of acute ankle-specific pain during one working day can end quickly with little effect on gait or can expand into a persistent gait problem with consistent prior expectation of chronic body-wide pain [51]. In particular, propensity to experience pain can be affected by personality type [52], hardness levels [53], fascia condition [54], and body memory [55]. As illustrated in Figure 8, this example illustrates the potential for internal models to go from being aligned with the external world (green interface) to becoming maladapted to the external world (red interface) [56, 57].

Two different embodied models of self in the world	Psychomotor predictive processing		
	Top-down Prior expectations consistency	Bottom-up Sensory signals strength	Intermediate Attention-based modulation
Low pain propensity embodied self-model	Specific acute pain only during work continued briefly after initial pain surprise	Directly related to physical impact during specific work activity	Attention elsewhere Forgetting
High pain propensity embodied self-model	Eventually, chronic body-wide pain throughout daily life irrespective of type of movement or whether there is any movement	Eventually, strength of bottom-up signals is not directly related to particular movements or to lack of movements	Eventually, hypervigilance to body-wide pain

Table 1: Different embodied internal models leading to different interface experiences

This example also illustrates that human internal models do not necessarily maintain continuous synchrony with the external world. Rather, there is a trade-off between accuracy of predictions based on prior expectations from models and the complexity of models. Individuals and organizations do need to fit with the world in which they intend to survive. In other words, they do need to have ecological fitness. However, sensory input is inherently noisy, and continually increasing the complexity of an internal model in response to every unexpected small sensory input would lead to over-fitting. In other words, the complexity of an internal model could increase through unknowingly taking noise to be underlying model structure. Conversely, under-fitting can occur when an agent's internal model does not adequately resemble the environment in which an agent intends to survive. For example, a linear internal model will tend to have poor predictive performance when an agent is trying to survive in a non-linear environment [58, 59, 60]. Thus, there needs to be a balance between model complexity and prediction accuracy. This need is equally relevant to individuals' embodied models and organizations' documented models. It is in this context that AML outputs, such as representations of the models produced in Section 3 will be interpreted.

4.3 Gait feature analyses as boundary objects

Both individuals and organizations can balance complexity and accuracy by simplifying their internal models to suit their biases and avoiding any potential source of sensory input from the external world

that might not support their biases [61]. For individuals this can lead to motivated cognition [38] and wishful seeing [39]. For organizations this can lead to lock-ins [40] and path dependency [41]. All of which can lead to information, such as the tree in Figure 7, being interpreted to serve explanations that support preconceptions [42] and confirm biases [43]. This very important issue of human interpretation to support preconceptions and biases is often overlooked in XAI literature that focuses on the technical challenges of obtaining and presenting information from machine learning models that can be interpreted by humans. In particular, there is little consideration of interpretable information from machine learning models as boundary objects between internal models and external states. As summarized in Figure 9, boundary objects are situated within interfaces between the boundaries of internal models and the external world. Boundary objects have different meanings in different social worlds but nonetheless can be meaningful to more than one social world. Boundary objects include written words, spatial diagrams, and any other types of information that can facilitate development and maintenance of coherence across intersecting world models [62, 63].

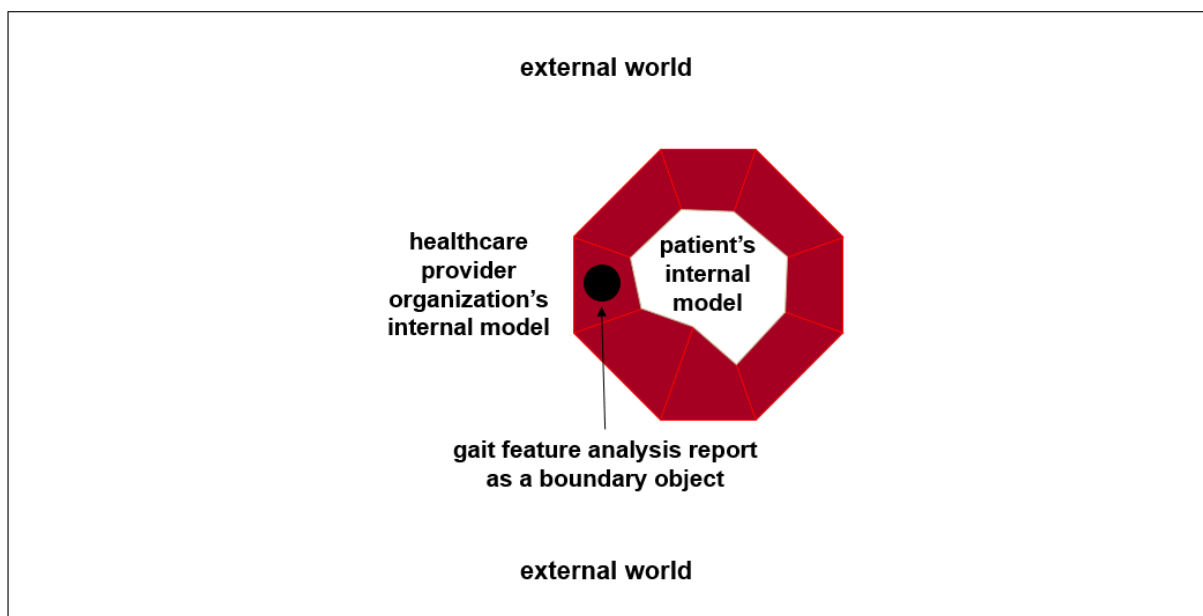


Figure 9: Gait feature analysis report as a boundary object

For example, gait feature analysis reports can be situated within the interface between the embodied internal model of a person suffering from a functional disorder accompanied with kinesiophobia, and a healthcare provider organization's documented internal model. To the person suffering from a functional disorder, the healthcare provider organization is in the external world. Gait analysis reports can contribute to evaluation of patients' complex disorder status, treatment, rehabilitation and recovery [1]. In practical terms, this can involve important outcomes such as preventing wheelchair dependency [64]. On the other hand, gait analysis reports could contribute to phobic avoidance and affective disorders that can perpetuate functional disorders [13], rather than contribute to a care pathway for patient recovery [65, 66].

4.4 AML-enabled gait feature analyses as boundary objects

As explained above, it is important that interpretable information from AML models, such as that shown in Figure 7, does not facilitate explanations that serve only to further entrench preconceptions. In relation

to functional disorders, it is possible for patients and healthcare providers to have strong preconceptions that are in opposition to each other.

From the point-of-view of healthcare providers, lack of complete medical explanation for functional disorders can lead to concerns that they are actually factitious disorders or malingering [67]. Factitious disorder, which has also been called Munchausen Syndrome, involves people behaving as if they have illnesses by deliberately producing, feigning, or exaggerating symptoms [68]. This is different to malingering that involves deliberate effort to simulate illness in order to get out of obligations and/or to obtain benefits [69]. Accordingly, there can be stigma against patients with functional disorders that presents obstacles to diagnosis and treatment. It has been argued that symptoms can be misunderstood or dismissed because of stigma. Moreover, it has been argued that stigma results in poor clinical management involving prolonged use of healthcare resources, and exacerbates suffering of patients [70].

Apropos, patients can have deeply entrenched preconceptions that can lead them to not favour some diagnoses and treatment recommendations. For example, patients who have kinesiophobia can be unreceptive to diagnosis and treatments that involve increasing their range of physical movement. This is because people who are suffering from kinesiophobia are afraid of moving. Specifically, people who are suffering from kinesiophobia have the deeply entrenched preconception that movement will cause pain. However, fear of movement can lead to increased disability and heightening of overall pain experience pain [49, 71]. Thus, kinesiophobia needs to be addressed. Yet, patients who do not agree with a diagnosis are less likely to respond to treatment [72]. Thus, there is an ethical dilemma concerning the allocation of scarce healthcare resources unless there is agreement about diagnosis and treatment recommendation.

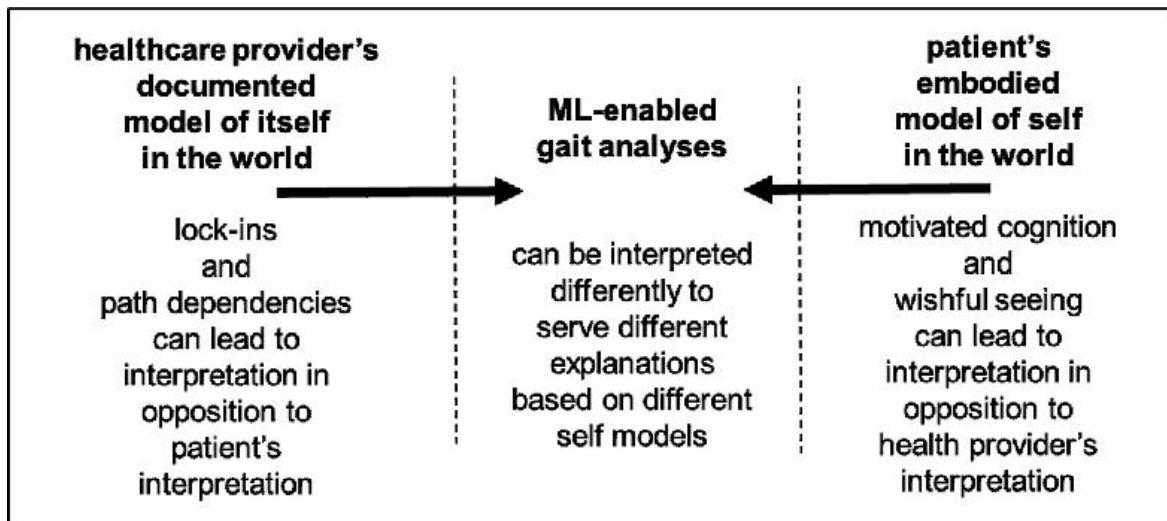


Figure 10: Potential for opposing interpretations from the same gait analysis

Accordingly, there are two fundamental questions to be addressed in evaluation of AML gait analyses. Is AML gait analysis at least equal to other machine learning (ML) gait analyses? Do AML gait analyses have more potential than other ML gait analyses to provide boundary object information that can support interpretations that serve to bridge opposing preconceptions?

With regard to the first question, our preliminary findings reported in Section 3.4 indicate that AML is a very promising approach for gait analysis. However, Whether or not AML can provide boundary objects that overcome the interpretability challenge summarized in Figure 10 remains unclear. The three characteristics of AML outlined in section 2.4 offer some potential for the co-development of boundary objects between healthcare providers and functional disorder patients. In particular, AML-models have

a persistently simple structure; top-down rules in the atoms in an initial AML-DL model are maintained throughout training; and outputs from AML-DL models can be traced back to human-defined policies. Accordingly, the co-development of AML-DL models by patients and healthcare providers would be consistent with the trend towards increased patient engagement. That is, enhancing patients ability to fully participate in healthcare, helping them to be equipped, enabled, and empowered regarding their own health. Patient engagement has been repeatedly linked to better health outcomes [73, 74]. Exploring this potential for AML is a direction for further research and development work, as can be seen from the tree in Figure 7, much improvement is required.

5 Conclusions

In this document, Deliverable D5.5 - AML based world models - Revised, we have reported findings from our continued exploratory research into the potential of Algebraic Machine Learning (AML) to support ethical world models. In doing so we have continued exploratory research started in D5.4. in addressing interfaces between individuals' embodied models of themselves in the world and organizations' documented models of themselves in the world. Here, the interface between individuals' embodied models and organizations' documented models is AML-supported analyses of gait features captured with cameras and/or sensors. In particular, our research reported in this Deliverable, 5.5, explored the potential of AML improve upon the explainability of AI-supported analyses of gait features.

Our exploratory research indicates that AML may have some potential to improve classification based on gait features and the explainability of gait feature analyses because of its simple three level structure and conformance to human-based constraints specified in human-defined policies that can be written in AML-DL. Our research also indicates that AML at this point may have limitations regarding overfitting in the absence of a validation set and size of the learned models. Moreover, in future work we will undertake research and development work to improve the interpretability of AML outputs as boundary objects.

This deliverable will be followed by D5.8 AML based world models - Final. In addition, Work Package 5 will deliver D5.6 Embedding of ethics and cultural concerns into AML based world models - Initial, and D5.7 Embedding of ethics and cultural concerns into AML based world models - Final. In those future deliverables, we shall report our further research into the potential of Algebraic Machine Learning (AML) to improve interactive human-centric machine learning systems in complex settings.

References

- [1] Chandra Prakash, Rajesh Kumar, and Namita Mittal. “Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges”. In: *Artificial Intelligence Review* 49.1 (2018), pp. 1–40.
- [2] MD Jacquelin Perry. “Gait analysis: normal and pathological function”. In: *New Jersey: SLACK* (2010).
- [3] Robert T Collins, Ralph Gross, and Jianbo Shi. “Silhouette-based human identification from body shape and gait”. In: *Proceedings of fifth IEEE international conference on automatic face gesture recognition*. IEEE. 2002, pp. 366–371.
- [4] Gabor Barton et al. “Gait quality assessment using self-organising artificial neural networks”. In: *Gait & posture* 25.3 (2007), pp. 374–379.
- [5] Joana Figueiredo, Cristina P Santos, and Juan C Moreno. “Automatic recognition of gait patterns in human motor disorders using machine learning: A review”. In: *Medical engineering & physics* 53 (2018), pp. 1–12.
- [6] Preeti Khera and Neelesh Kumar. “Role of machine learning in gait analysis: a review”. In: *Journal of Medical Engineering & Technology* 44.8 (2020), pp. 441–467.
- [7] Roman Schniepp, Ken Möhwald, and Max Wuehr. “Clinical and automated gait analysis in patients with vestibular, cerebellar, and functional gait disorders: perspectives and limitations”. In: *Journal of neurology* 266.1 (2019), pp. 118–122.
- [8] Djordje Slijepcevic et al. “Automatic classification of functional gait disorders”. In: *IEEE journal of biomedical and health informatics* 22.5 (2017), pp. 1653–1661.
- [9] Bogdan Pogorelc, Zoran Bosnić, and Matjaž Gams. “Automatic recognition of gait-related health problems in the elderly using machine learning”. In: *Multimedia tools and applications* 58.2 (2012), pp. 333–354.
- [10] Mingjing Yang et al. “A machine learning approach to assessing gait patterns for complex regional pain syndrome”. In: *Medical engineering & physics* 34.6 (2012), pp. 740–746.
- [11] Jeffrey M Hausdorff et al. “Gait unsteadiness and fall risk in two affective disorders: a preliminary study”. In: *BMC psychiatry* 4.1 (2004), pp. 1–7.
- [12] Stoyan Popkirov et al. “Complex regional pain syndrome and functional neurological disorders—time for reconciliation”. In: *Journal of Neurology, Neurosurgery & Psychiatry* 90.5 (2019), pp. 608–614.
- [13] Alberto J Espay et al. “Current concepts in diagnosis and treatment of functional neurological disorders”. In: *JAMA neurology* 75.9 (2018), pp. 1132–1141.
- [14] Kati Thieme, Dennis C Turk, and Herta Flor. “Comorbid depression and anxiety in fibromyalgia syndrome: relationship to somatic and psychosocial variables”. In: *Psychosomatic medicine* 66.6 (2004), pp. 837–844.
- [15] Nan Zhao et al. “See your mental state from your walk: Recognizing anxiety and depression through Kinect-recorded gait data”. In: *PLoS one* 14.5 (2019), e0216591.
- [16] Djordje Slijepcevic et al. “On the Explanation of Machine Learning Predictions in Clinical Gait Analysis”. In: *arXiv preprint arXiv:1912.07737* (2019).
- [17] Fernando Martin-Maroto and Gonzalo G de Polavieja. “Algebraic Machine Learning”. In: *arXiv preprint arXiv:1803.05252* (2018).
- [18] Cynthia Rudin. “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead”. In: *Nature Machine Intelligence* 1.5 (2019), pp. 206–215.
- [19] Scott M Lundberg and Su-In Lee. “A unified approach to interpreting model predictions”. In: *Advances in neural information processing systems* 30 (2017).
- [20] Christoph Molnar. *Interpretable machine learning*. Feb. 2022. URL: <https://christophm.github.io/interpretable-ml-book/index.html>.
- [21] Yameng Wang et al. “Detecting depression through gait data: examining the contribution of gait features in recognizing depression”. In: *Frontiers in psychiatry* 12 (2021).

- [22] Wei-Yin Loh. “Classification and regression trees”. In: *Wiley interdisciplinary reviews: data mining and knowledge discovery* 1.1 (2011), pp. 14–23.
- [23] Thomas Metzinger. “Empirical perspectives from the self-model theory of subjectivity: a brief summary with examples”. In: *Progress in brain research* 168 (2007), pp. 215–278.
- [24] Anil K Seth and Manos Tsakiris. “Being a beast machine: the somatic basis of selfhood”. In: *Trends in cognitive sciences* 22.11 (2018), pp. 969–981.
- [25] Micah Allen and M Tsakiris. “The body as first prior: Interoceptive predictive processing and the primacy”. In: *The interoceptive mind: From homeostasis to awareness* 27 (2018).
- [26] Zoë C Franklin et al. “Personality type influences attentional bias in individuals with chronic back pain”. In: *PloS one* 11.1 (2016), e0147035.
- [27] Enrique Merino-Tejedor et al. “Hardiness as a moderator variable between the Big-Five Model and work effort”. In: *Personality and Individual Differences* 85 (2015), pp. 105–110.
- [28] Carla Stecco. *Functional atlas of the human fascial system*. Elsevier Health Sciences, 2014.
- [29] Giuseppe Riva. “The neuroscience of body memory: From the self through the space to the others”. In: *Cortex* 104 (2018), pp. 241–260.
- [30] Stephen Fox. “Practical implications from differentiating between Pearl blankets and Friston blankets”. In: *Behavioral and Brain Sciences* (2022). accepted, but yet published.
- [31] Nick Haslam. “Unicorns, snarks, and personality types: A review of the first 102 taxometric studies of personality”. In: *Australian Journal of Psychology* 71.1 (2019), pp. 39–49.
- [32] Paolo Tozzi. *Does fascia hold memories?* 2014.
- [33] Jay J Schlickman. *ISO 9001: 2000 quality management system design*. Artech House, 2003.
- [34] Muhammad Adnan Khurshid, Muslim Amin, and Wan Khairuzzaman Wan Ismail. “Total quality and socially responsible management (TQSR-M): an integrated conceptual framework”. In: *Benchmarking: An International Journal* (2018).
- [35] Pervaiz K Ahmed and Silke Machold. “The quality and ethics connection: toward virtuous organizations”. In: *Total Quality Management & Business Excellence* 15.4 (2004), pp. 527–545.
- [36] Irene Tracey. “Getting the pain you expect: mechanisms of placebo, nocebo and reappraisal effects in humans”. In: *Nature medicine* 16.11 (2010), pp. 1277–1283.
- [37] John Argenti. *Practical corporate planning*. Routledge, 2018.
- [38] Matthew S Nurse and Will J Grant. “I’ll see it when I believe it: Motivated numeracy in perceptions of climate change risk”. In: *Environmental Communication* 14.2 (2020), pp. 184–201.
- [39] David Dunning and Emily Balcetis. “Wishful seeing: How preferences shape visual perception”. In: *Current directions in psychological science* 22.1 (2013), pp. 33–37.
- [40] Allan I Teger. *Too Much Invested to Quit: Pergamon General Psychology Series*. Vol. 83. Elsevier, 2017.
- [41] Jörg Sydow, Georg Schreyögg, and Jochen Koch. “Organizational path dependence: Opening the black box”. In: *Academy of management review* 34.4 (2009), pp. 689–709.
- [42] Philip E Tetlock. “Theory-driven reasoning about plausible pasts and probable futures in world politics: are we prisoners of our preconceptions?” In: *American Journal of Political Science* (1999), pp. 335–366.
- [43] Rüdiger Pohl and Rüdiger F Pohl. *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory*. Psychology Press, 2004.
- [44] Thomas Deroche et al. “Athletes’ inclination to play through pain: a coping perspective”. In: *Anxiety, Stress & Coping* 24.5 (2011), pp. 579–587.
- [45] Andrew Flood et al. “Increased conditioned pain modulation in athletes”. In: *Journal of sports sciences* 35.11 (2017), pp. 1066–1072.
- [46] Stephen Fox et al. “Aligning human psychomotor characteristics with robots, exoskeletons and augmented reality”. In: *Robotics and Computer-Integrated Manufacturing* 63 (2020), p. 101922.
- [47] Marie-Héloïse Bardel et al. “The role of athletes’ pain-related anxiety in pain-related attentional processes”. In: *Anxiety, Stress & Coping* 26.5 (2013), pp. 573–583.

- [48] Dean A Tripp et al. “Fear of reinjury, negative affect, and catastrophizing predicting return to sport in recreational athletes with anterior cruciate ligament injuries at 1 year postsurgery.” In: *Rehabilitation psychology* 52.1 (2007), p. 74.
- [49] Jeffrey Roelofs et al. “The Tampa Scale for Kinesiophobia: further examination of psychometric properties in patients with chronic low back pain and fibromyalgia”. In: *European Journal of Pain* 8.5 (2004), pp. 495–502.
- [50] Fatemeh Heidari, Mahdi Afshari, and Mahmood Moosazadeh. “Prevalence of fibromyalgia in general population and patients, a systematic review and meta-analysis”. In: *Rheumatology international* 37.9 (2017), pp. 1527–1539.
- [51] Stephen Fox. “Psychomotor predictive processing”. In: *Entropy* 23.7 (2021), p. 806.
- [52] Angelina R Sutin et al. “The prospective association between personality traits and persistent pain and opioid medication use”. In: *Journal of psychosomatic research* 123 (2019), p. 109721.
- [53] Mike Soderstrom et al. “The relationship of hardiness, coping strategies, and perceived stress to symptoms of illness”. In: *Journal of behavioral medicine* 23.3 (2000), pp. 311–328.
- [54] Piero G Pavan et al. “Painful connections: densification versus fibrosis of fascia”. In: *Current pain and headache reports* 18.8 (2014), pp. 1–8.
- [55] Tess Burton. “Painful memories: Chronic pain as a form of re-membering”. In: *Memory Studies* 4.1 (2011), pp. 23–32.
- [56] Achim Peters, Bruce S McEwen, and Karl Friston. “Uncertainty and stress: Why it causes diseases and how it is mastered by the brain”. In: *Progress in neurobiology* 156 (2017), pp. 164–188.
- [57] Karl Sallin et al. “Resignation syndrome: catatonia? Culture-bound?” In: *Frontiers in Behavioral Neuroscience* (2016), p. 7.
- [58] Brian S Everitt and Anders Skrondal. “The Cambridge dictionary of statistics”. In: (2010).
- [59] D Anderson and K Burnham. “Model selection and multi-model inference”. In: *Second. NY: Springer-Verlag* 63.2020 (2004), p. 10.
- [60] Roger C Conant and W Ross Ashby. “Every good regulator of a system must be a model of that system”. In: *International journal of systems science* 1.2 (1970), pp. 89–97.
- [61] Arie W Kruglanski, Katarzyna Jasko, and Karl Friston. “All thinking is ‘wishful’ thinking”. In: *Trends in Cognitive Sciences* 24.6 (2020), pp. 413–424.
- [62] Susan Leigh Star and James R Griesemer. “Institutional ecology, translations’ and boundary objects: Amateurs and professionals in Berkeley’s Museum of Vertebrate Zoology, 1907–39”. In: *Social studies of science* 19.3 (1989), pp. 387–420.
- [63] Geoffrey C Bowker and Susan Leigh Star. *Sorting things out: Classification and its consequences*. MIT press, 2000.
- [64] José Fidel Baizabal-Carvallo, Marlene Alonso-Juarez, and Joseph Jankovic. “Functional gait disorders, clinical phenomenology, and classification”. In: *Neurological Sciences* 41.4 (2020), pp. 911–915.
- [65] Davina Allen. “From boundary concept to boundary object: the practice and politics of care pathway development”. In: *Social Science & Medicine* 69.3 (2009), pp. 354–361.
- [66] Jon Stone. “Functional neurological disorders: the neurological assessment as treatment”. In: *Practical Neurology* 16.1 (2016), pp. 7–17.
- [67] C Bass and P Halligan. “Factitious disorders and malingering in relation to functional neurologic disorders”. In: *Handbook of clinical neurology* 139 (2016), pp. 509–520.
- [68] Xavier F Jimenez et al. “Clinical, demographic, psychological, and behavioral features of factitious disorder: A retrospective analysis”. In: *General Hospital Psychiatry* 62 (2020), pp. 93–95.
- [69] Christopher Bass and Derick T Wade. “Malingering and factitious disorder”. In: *Practical Neurology* 19.2 (2019), pp. 96–105.
- [70] Katherine E MacDuffie et al. “Stigma and functional neurological disorder: a research agenda targeting the clinical encounter”. In: *CNS spectrums* 26.6 (2021), pp. 587–592.

- [71] Jeffrey Roelofs et al. “Fear of movement and (re) injury in chronic musculoskeletal pain: Evidence for an invariant two-factor model of the Tampa Scale for Kinesiophobia across pain diagnoses and Dutch, Swedish, and Canadian samples”. In: *Pain* 131.1-2 (2007), pp. 181–190.
- [72] Sarah C Lidstone, Lindsey MacGillivray, and Anthony E Lang. “Integrated therapy for functional movement disorders: time for a change”. In: *Movement disorders clinical practice* 7.2 (2020), pp. 169–174.
- [73] Faith Birnbaum et al. “Patient engagement and the design of digital health”. In: *Academic emergency medicine: official journal of the Society for Academic Emergency Medicine* 22.6 (2015), p. 754.
- [74] Judith H Hibbard et al. “Do increases in patient activation result in improved self-management behaviors?” In: *Health services research* 42.4 (2007), pp. 1443–1463.