Fuzzy Rule Based Ensemble for Classification of Gait Patterns in Cerebral Palsy Patients

Rodrigo B. Ventura, Filipe M.P. Santos, Susana M. Vieira, and João M.C. Sousa

IDMEC, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, 1049-001, Portugal

Abstract. Applications of data-based modeling tools in medical diagnosis are one of the most promising trends in the healthcare field, which has been expedited by the increasingly common usage of various measuring devices and monitoring systems. This paper addresses one such application, namely the classification of gait patterns in cerebral palsy patients, using various kinetic and kinematic measurements describing the walking patterns. In order to address the data fusion challenge resultant from the multi-sensor environment, while also meeting the desirable explainability properties of medical models, this paper proposes a multi-view ensemble approach based on multiple fuzzy rule based classifiers, each one assigned to a single sensor and its measurements. The proposed approach is tested on four different problems with different gait pattern anomalies, using data from real subjects obtained using state-of-art biomechanical methods and standard clinical procedures.

Keywords: Cerebral Palsy · Gait Patterns · Fuzzy Systems · Multi-view Ensemble · Data Fusion.

1 Introduction

Cerebral palsy refers to a group of permanent disorders in the development of movement and posture caused by non-progressive disturbances in the developing fetal or infant brain [1]. The most common type of cerebral palsy is known as spastic diplegia, which affects the motor cortex of the brain and controls voluntary movement [2]. Spastic diplegia is characterized by muscle tightness in the legs, hips and pelvis, thus having negative effects in the walking functions of the patient.

Moreover, these effects vary by individual, and must be accurately diagnosed for various medical interventions. Therefore, standardized categories were created for the various types of walking function anomalies. These categories are usually defined by quantitative descriptions of the anomalous walking patterns, known as gait patterns [3].

This paper presents a data-driven approach to diagnostic models for patients suffering from spastic diplegia, using real measurements from multiple test subjects. The dataset includes multiple samples, each one consisting of multiple sensor measurements structured as time series. Each sensor measures different parameters of the gait patterns recorded for the test subjects [4]. The diagnostic models must predict the gait pattern anomaly (labeled by a medical expert) using the various sensor readings [5]. Thus, the described problem not only presents the already expected challenges often found in medical datasets [6], but also requires data fusion of the various data sources.

In order to adequately address the multi-source dataset structure, this work proposes a multi-view ensemble approach [8] with multiple base models, each one assigned to a sensor [7]. Moreover, applications of data-based modeling in medical applications, such as the one addressed in this work, increasingly require interpretable [9] models in lieu of traditional black-box modeling tools. Thus, we also address this requirement in the proposed approach, by structuring the ensemble base models as fuzzy inference systems.

2 Problem Description

The problem addressed in this work is the data-based identification of medical diagnostic models for detection of different gait pattern anomalies. The dataset used in this paper was obtained for a group of test subjects in a controlled laboratory environment. Each subject performed multiple gait cycles, each one corresponding to an unique sample, consisting of the recorded sensor measurements (features), and the gait pattern (class label) detected by the medical expert supervising the tests.

The described procedure was conducted for a group of 51 subjects, consisting of a control group with 25 healthy children, and a patient group with 26 children diagnosed with spastic diplegia. The patient group is further divided in four different groups depending on the gait pattern anomaly affecting each individual. The four groups correspond to Apparent Equinus, Crouch Gait, Jump Gait, and True Equinus, all of which are commonly found in spastic diplegia patients.

Subjects belonging to the patient group can suffer from different gait pattern anomalies in each leg. There are also some cases in which the measurements are only available for one of the legs, either as a consequence of sensor failure, or lack of a clear gait pattern label provided by the medical expert. Since each leg has its own independent set of measurements, we consider each leg as a separate entry in the dataset, regardless of the subject. Therefore, each sample corresponds to a set of gait cycle measurements, recorded for a single leg, resulting in the dataset summarized in Table 1.

Regarding the laboratory setup for recording the gait cycles, a total of 15 motion tracking cameras and 3 force platforms were used. Regarding the cameras, these are used to localize reflective markers attached to the subject. This data is then processed to generate a three-dimensional link-segment model of the subject's body, which is then used for estimating the body joint angles (kinematic features). Regarding the force platforms, they provide measurements of the ground reaction forces during a walking test. These force measurements can

Table 1. Sample availability and respective class distribution.

Group	Gait Pattern	Available Samples	
		Legs	**Gait Cycles**
Control	Normal	50	183
Patient	Apparent Equinus	8	19
	Crouch Gait	8	23
	Jump Gait	10	29
	True Equinus	9	27

then be used to estimate the body joint moments (kinetic features) by solving the resultant inverse dynamics problem .

The aforementioned procedure results in a dataset structure consisting of 21 independent sensor measurements. Kinematic measurements consist of distinct joint angles measured in degrees. These measurements are obtained for 4 body joints (ankle, hip, knee and pelvis) in the X, Y and Z planes, resulting in 12 unique features. Kinetic measurements consist of distinct joint moments, and are measured for 3 body joints (ankle, hip, knee) in the X, Y and Z planes, resulting in 9 unique features. Each one of the unique 21 measurements is structured as a time series with 101 data points, resulting in a total of 2121 data points for each gait cycle recorded.

3 Methodology

This section presents the proposed approach and its elements, describing the multi-view ensemble architecture (section 3.1) and the fuzzy rule based structure of the base models (section 3.2).

3.1 Multi-View Ensemble Architecture

As discussed before, information fusion and multi-view approaches include a diverse set of problem formulations and dataset structures, resulting in different data-based modeling tasks with varying levels of complexity. Attending to the characteristics of problem addressed in this paper, the proposed approach is adequate to classification tasks on datasets already structured as multiple data views, not requiring any procedure to obtain the feature subsets.

Consider a multi-view dataset $\mathcal{D}$ with n samples, m features, and v data views, such that $\mathcal{D} = \left\{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots \mathbf{X}^{(v)}\right\}$, where $\mathbf{X}^{(i)} = \left[\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \ldots, \mathbf{x}_n^{(i)}\right] \in \mathcal{R}^{n \times m_i}$ represents the i-th data view and m_i its number of features. Therefore, each data view has the same number of samples, but an unique subset of features, such that $m = \sum_{i=1}^{v} m_i$ is the total number of features. For a classification problem with c classes, we define the class labels as $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_n\} \in \mathcal{R}^{n \times c}$, where $\mathbf{y}_k = [y_{k1}, y_{k2}, \ldots, y_{kc}]$ defines the label of the k-th sample, with $y_{kj} = 1$ if the sample belongs to the j-th class, and $y_{kj} = 0$ otherwise.

Having defined the generic problem for a multi-view dataset, we can now start defining the proposed ensemble structure. First, we define the set of base classifiers as $\mathcal{B} = \left\{ f^{(1)}\left(\theta^1, \mathbf{x}_k^{(1)}\right), f^{(2)}\left(\theta^2, \mathbf{x}_k^{(2)}\right), \ldots, f^{(v)}\left(\theta^v, \mathbf{x}_k^{(v)}\right) \right\}$ where $f^{(i)}\left(\theta^i, \mathbf{x}_k^{(i)}\right)$ represents the i-th base classifier as a mapping from the k-th sample $\mathbf{x}_k^{(i)}$ to the respective class prediction $\mathbf{y}_k^{(i)}$, using the estimated model parameters θ^i. Therefore, the first layer in the ensemble is fully defined by estimating the parameters corresponding to each model. Attending to the parallel ensemble structure, each training procedure is fully independent and can be performed simultaneously. After the base model parameters are estimated, the first ensemble layer is fully defined in its final state.

Before proceeding to the aggregation step, we must first define a single layer output $\hat{\mathbf{w}}$ from the v class predictions as $\hat{\mathbf{w}} = \left[\hat{\mathbf{y}}^{(1)}, \hat{\mathbf{y}}^{(2)}, \ldots, \hat{\mathbf{y}}^{(v)}\right] \in \mathcal{R}^{v \times c}$, with $\hat{w}_{ij}$ defining the i-th classifier prediction for the j-th class. The first ensemble layer can therefore be defined as $\hat{\mathbf{w}} = \mathcal{B}(\mathbf{x})$ defining a mapping from sample $\mathbf{x}$ to the base model prediction. All the previous definitions are valid for hard classifiers predicting the class label, as well as soft classifiers predicting the class probabilities. For hard classifiers, the aforementioned encoding results in a probability of 1.0 for the predicted class, and 0.0 for other classes.

The final step in the proposed ensemble consists of combining the base model predictions in order to obtain a single output corresponding to the ensemble class prediction. This step is known as aggregation and presents an important design choice in ensemble models with noticeable impact on predictive performance. Well known strategies include simpler non-parametric methods, such as different voting rules, as well as more complex parametric methods, such as stacking. Attending to the interest in studying the impact of different aggregation strategies on the predictive performance, we propose three different methods for this step, and will later discuss the obtained results.

Regardless of the chosen method, the aggregation step can be defined as $\hat{\mathbf{y}} = \mathcal{S}(\hat{\mathbf{w}})$ with $\hat{\mathbf{y}} \in \mathcal{R}^c$. Starting with simpler non-parametric methods, we propose a simple voting strategy defining $\hat{y_j} = \sum_{i=1}^{v} \hat{w_{ij}}$ as the total voting score for the j-th class, and predicting the class with maximum score. This voting strategy is valid for hard and soft predictions, with the latter case generally improving the resultant predictive performance. Both cases are studied and compared in the results discussion. In order to generalize the aggregation procedure, we also propose a data-based modeling approach for the estimation of $\mathcal{S}(\hat{\mathbf{w}})$ as a separate ensemble base model. This parametric approach defines a classifier mapping base predictions to the final class prediction, allowing for more complex class decision rules. Such approaches are known as stacking and are also well known and commonly used in multi-view ensembles.

3.2 Fuzzy Rule Based Classifiers

In this work, the ALMMo-0 classifier [10] was chosen for identification of all the ensemble base classifiers that were used for testing the proposed approach, as will be discussed in section 4. This choice was primarily motivated by the simplicity and performance shown in the literature. Moreover, recent publications have shown the adequacy of ALMMo-0 classifiers for medical diagnosis applications [11], as well as the building block for complex ensemble architectures and deep fuzzy approaches [12].

As part of the Empirical Data Analytics (EDA) [13] framework ALMMo-0 classifiers are based on zero-order AnYa type fuzzy rules [14], presenting a simplified antecedent structure and training algorithm. The model structure and respective parameters are recursively updated in a non-iterative and feed-forward way, on a sample-by-sample basis, forming data clouds which define the antecedents.

Regarding the specific problem addressed in this work, the primary reason for choosing the ALMMo-0 is its robust performance when compared to similarly complex methods, particularly in challenging classification problems and highly imbalanced datasets. One key aspect supporting such performances is the rule structure, and the separate rule (cloud) identification process for each class. This structure differs from the structure used in traditional fuzzy models, such as Takagi-Sugeno [15] and Mamdani fuzzy systems [16].

In the remainder of this section, we describe the rule structure and its consequences regarding predictive performance. Attending to the scope of this paper, and also for the sake of brevity, the cloud identification algorithm is not described, as it is the same proposed in [10]. The general architecture consists of multiple parallel fuzzy rule based sub-models, one for each class. Moreover, each rule corresponds to a data cloud, which defines its antecedents. In the case of ALMMo-0 classifiers, rules are completely defined by the antecedent parameters, since zero-order AnYa-type fuzzy rules have non-parametric consequents. The rules are structured as shown in (1), where θ_k^c represents the focal point describing the k-th cloud from the c-th class sub-model, and λ_k^c is defined as $\lambda_k^c = exp\left(-\frac{1}{2}\|\mathbf{x} - \theta_k^c\|^2\right)$ and represents the respective activation score.

$$\mathbf{IF}\ \ \mathbf{x} \sim \theta_k^c\ \ \mathbf{THEN}\ \hat{y}^c \ = \ \lambda_k^c \tag{1}$$

Each one of these class sub-models has its unique set of fuzzy rules, since these are updated using only the respective class samples. Therefore, an independent fuzzy rule set is created for each class, thus avoiding common problems related to the existence of majority class outliers and noisy samples, generally improving model performance, particularly in highly imbalanced datasets.

When classifying an unknown data sample, each one of the class sub-models receives the sample as an input to each one of the rules in its rule set. Then, each sub-model returns the maximum activation score found in its rule set, resulting in the rule structure shown in (2), where λ_*^c is the maximum confidence score and R_c is the number of rules in the c-th class sub-model.

$$\mathbf{IF}\ \ (\mathbf{x} \sim \theta_1^c)\ \ \mathbf{OR}\,(\mathbf{x} \sim \theta_2^c)\ \ \mathbf{OR}\ ...\ \mathbf{OR}\left(\mathbf{x} \sim \theta_{R_c}^c\right)\ \ \mathbf{THEN}\ \hat{y}^c \ = \ \lambda_*^c \tag{2}$$

In the original paper, the maximum scores of each sub-model are then compared, and the largest score assigns its class label to the sample, using a winner-takes-all strategy. In this paper, we propose a small change to this last step, defining a normalization layer for the confidence scores. We define the model output as $\hat{\mathbf{y}} = \left[y^1, y^2, \dots y^c\right]$, where $y^i = \lambda_*^i / \sum_{j=1}^{c} \lambda_*^j$ is the i-th class probability. This allows the ALMMo-0 to be used with soft voting methods, as discussed in 3.1, and does not affect the model performance of the original algorithm.

4 Results

In order to more thoroughly assess the validity of the proposed approach, we define four different problems from the dataset presented in section 2, each one defined as the one-vs-rest binary problem for each gait anomaly. Following this procedure, we get the binary classification problems summarized in Table 2.

Table 2. Sample distribution and respective class distribution for each one of the binary problems obtained from one-vs-rest decomposition.

Problem	**Sample Distribution**		**Class Imbalance**
(Target vs Rest)	**Target**	**Rest**	**(%)**
Apparent Equinus vs Rest	19	262	6.8
Crouch Gait vs Rest	23	258	8.2
Jump Gait vs Rest	29	252	10.3
True Equinus vs Rest	27	254	9.6

For each one of the problems 5-fold cross validation was used. The obtained results are presented for each problem separately, in Table 3, Table 4, Table 5 and Table 6, respectively. The predictive performance is evaluated using well known classification metrics, including Accuracy, Recall, Precision, F1-score, Cohen's Kappa Coefficient (Kappa) and Matthews Correlation Coefficient (MCC).

Moreover, in order to facilitate the ensemble performance analysis, the tables are divided row-wise in two groups of base model results, and one group of ensemble model results. Starting from the top, the first 12 rows show the results for base models trained on angle measurements for distinct body joints, which include the ankles (A-An), the hips (A-Hp), the knees (A-Kn), and pelvis (A-Pe). The 9 middle rows of each table show the results for base models trained on moment measurements for distinct body joints, which include the ankles (M-An), the hips (M-Hp), and the knees (M-Kn). Finally, the last 3 rows in each table show the overall ensemble results using 3 distinct aggregation methods for the base classifier predictions. These include hard voting (HV), soft voting (SV), and stacking (ST). Values in bold correspond to the highest metric scores in each one the 3 groups, while underlined values mark the best overall.

Starting with Table 3 and the results for the apparent equinus detection, it is quite clear that the performance of the proposed ensemble was far from

Table 3. Performance metrics for the Apparent Equinus anomaly detection.

Model		Accuracy	Precision	Recall	F1-Score	Kappa	MCC
A-An	X	0.862	0.092	0.090	0.088	0.015	0.016
	Y	0.909	0.120	0.120	0.120	0.088	0.084
	Z	**0.929**	0.333	0.140	0.181	0.172	0.197
A-Hp	X	0.858	0.111	0.220	0.147	0.081	0.086
	Y	0.885	0.000	0.000	0.000	-0.048	-0.054
	Z	0.857	0.168	**0.320**	0.217	0.154	0.166
A-Kn	X	0.880	0.207	0.130	0.154	0.092	0.098
	Y	0.913	**0.524**	0.270	**0.330**	**0.289**	**0.320**
	Z	0.869	0.000	0.000	0.000	-0.055	-0.059
A-Pe	X	0.883	0.000	0.000	0.000	-0.045	-0.046
	Y	0.866	0.025	0.040	0.031	-0.033	-0.036
	Z	0.896	0.117	0.120	0.117	0.072	0.068
M-An	X	0.804	0.161	0.360	0.204	0.118	0.139
	Y	0.812	0.049	0.090	0.060	-0.029	-0.032
	Z	0.867	0.062	0.090	0.073	0.008	0.006
M-Hp	X	0.840	0.040	0.040	0.040	-0.042	-0.045
	Y	0.891	**0.500**	0.190	0.267	0.215	0.250
	Z	0.886	0.000	0.000	0.000	-0.047	-0.049
M-Kn	X	**0.904**	0.383	**0.370**	**0.307**	**0.266**	**0.301**
	Y	0.887	0.000	0.000	0.000	-0.045	-0.048
	Z	0.881	0.067	0.040	0.050	-0.002	-0.002
Ensemble	HV	**0.922**	0.000	0.000	0.000	0.000	0.000
	SV	**0.922**	0.000	0.000	0.000	0.000	0.000
	ST	**0.922**	0.000	0.000	0.000	0.000	0.000

acceptable, since it misclassified all the positive samples in all tests, regardless of the aggregation method. Observing the base model results, most outperform the ensembles in all metrics (except accuracy), as expected. Nevertheless, the precision and recall scores are generally quite low for all models, suggesting that the classification problem is rather challenging, regardless of the approach complexity of the modelling method. Despite the overall underperformance, it is also evident that base models trained on the Y axis knee joint angle (A-Kn-Y), as well as base models trained on the X axis knee joint moment (M-Kn-X), clearly outperform all other models, suggesting that these two features are by far the most informative regarding the apparent equinus gait anomaly.

Regarding the results for the crouch gait anomaly shown in Table 4, the ensemble approach with hard vote aggregation outperforms the other ensembles in all metrics. Moreover, it also outperforms all other models in terms of accuracy and precision. However, this suggests that the proposed ensemble (using hard voting) predicts a large number of false negatives, meaning that it often fails

Table 4. Performance metrics for the Crouch Gait anomaly detection.

Model		Accuracy	Precision	Recall	F1-Score	Kappa	MCC
A-An	X	0.907	0.496	0.593	0.532	0.481	0.489
	Y	0.896	0.463	0.640	0.533	0.476	0.487
	Z	0.921	0.587	0.640	0.610	0.566	0.568
A-Hp	X	**0.929**	0.622	**0.760**	**0.674**	**0.636**	**0.646**
	Y	0.919	0.597	0.640	0.598	0.554	0.566
	Z	0.904	0.458	0.373	0.393	0.344	0.355
A-Kn	X	0.816	0.260	0.467	0.328	0.234	0.250
	Y	0.928	**0.681**	0.580	0.593	0.556	0.575
	Z	0.865	0.177	0.273	0.209	0.149	0.152
A-Pe	X	0.764	0.000	0.000	0.000	-0.106	-0.119
	Y	0.892	0.461	0.460	0.437	0.379	0.392
	Z	0.841	0.102	0.113	0.107	0.022	0.021
M-An	X	0.877	0.457	0.447	0.375	0.321	0.359
	Y	**0.919**	**0.583**	0.440	0.480	0.441	0.456
	Z	0.891	0.397	0.347	0.365	0.307	0.310
M-Hp	X	0.846	0.120	0.147	0.130	0.049	0.049
	Y	0.873	0.442	0.467	0.424	0.358	0.374
	Z	0.904	0.467	0.207	0.267	0.232	0.262
M-Kn	X	0.903	0.542	**0.787**	**0.624**	**0.575**	**0.597**
	Y	0.892	0.251	0.287	0.265	0.222	0.222
	Z	0.876	0.327	0.527	0.390	0.336	0.356
Ensemble	HV	**0.935**	**0.800**	**0.460**	**0.565**	**0.533**	**0.566**
	SV	0.929	**0.800**	0.393	0.500	0.468	0.513
	ST	0.905	0.380	0.180	0.217	0.191	0.218

to detect the crouch gait anomaly (the low recall scores further support this conclusion). Moreover, these results show that hard voting can lead to better results than soft voting and stacking, as was discussed in section 3.1, despite being the simplest strategy. Regarding the base models, the best overall performances correspond to the classifiers trained on the X hip joint angle (A-Hp-X) and X knee joint moment (M-Kn-X) measurements, showing the highest recall, F1-Score, Kappa and MCC scores. Therefore, it is clear that once again the base models outperformed the ensembles.

Regarding the results in Table 5 for the jump gait anomaly detection, the proposed ensemble (using stacking as the aggregation method) outperforms all the remaining models in all metrics except the recall score. As such, these results show that the proposed ensemble approach can lead to more balanced classification performances than the individual base models, at the expense of a larger number of false negatives. Regarding the base classifiers, the best overall perfor-

Table 5. Performance metrics for the Jump Gait anomaly detection.

Model		Accuracy	Precision	Recall	F1-Score	Kappa	MCC
A-An	X	0.898	0.489	**0.720**	**0.563**	**0.510**	**0.534**
	Y	0.828	0.130	0.147	0.134	0.041	0.042
	Z	0.888	0.407	0.260	0.273	0.223	0.246
A-Hp	X	0.867	0.196	0.240	0.206	0.145	0.146
	Y	0.869	0.458	0.447	0.388	0.322	0.355
	Z	0.890	0.267	0.187	0.218	0.172	0.175
A-Kn	X	0.844	0.273	0.433	0.329	0.246	0.259
	Y	0.899	**0.529**	0.587	0.538	0.483	0.494
	Z	**0.913**	0.383	0.307	0.331	0.303	0.308
A-Pe	X	0.875	0.400	0.593	0.473	0.405	0.418
	Y	0.855	0.000	0.000	0.000	-0.049	-0.054
	Z	0.877	0.357	0.253	0.279	0.217	0.229
M-An	X	0.850	0.368	**0.827**	**0.506**	**0.433**	**0.485**
	Y	0.850	0.349	0.607	0.432	0.356	0.380
	Z	0.870	**0.447**	0.467	0.420	0.353	0.372
M-Hp	X	0.848	0.251	0.533	0.337	0.261	0.291
	Y	0.851	0.207	0.187	0.173	0.100	0.108
	Z	0.876	0.200	0.153	0.170	0.116	0.119
M-Kn	X	**0.886**	0.443	0.393	0.395	0.339	0.349
	Y	0.854	0.338	0.600	0.427	0.350	0.373
	Z	0.837	0.244	0.293	0.256	0.169	0.174
Ensemble	HV	0.921	0.467	0.273	0.337	0.319	0.336
	SV	0.918	0.600	0.153	0.238	0.226	0.288
	ST	**0.945**	**0.700**	**0.533**	**0.580**	**0.564**	**0.583**

mances were obtained for the X ankle joint angle (A-An-X) and X ankle joint moment (M-An-X).

Proceeding to Table 6 and the results for the true equinus pattern detection, it is clear that the stack ensemble approach once again outperforms the other aggregation strategies in all metrics. Nevertheless, the results shown once again that the base models still lead to the best overall classification performances. Specificaly, base classifiers trained on the X knee joint angle (A-Kn-X) and the X knee joint moment (M-Kn-X) features show the best overall performance.

5 Conclusions and Future Work

This paper proposes a data-based modeling approach to gait pattern detection in spastic diplegia patients. Validation of the presented approach was conducted using data from real subjects and following state-of-the-art clinical procedures.

Table 6. Performance metrics for the True Equinus anomaly detection.

Model		Accuracy	Precision	Recall	F1-Score	Kappa	MCC
A-An	X	0.841	0.147	0.124	0.134	0.058	0.053
	Y	0.826	0.117	0.062	0.080	-0.009	-0.008
	Z	0.886	0.419	0.286	0.329	0.276	0.284
A-Hp	X	0.852	0.407	**0.614**	0.480	0.398	0.417
	Y	0.814	0.172	0.229	0.184	0.084	0.092
	Z	0.847	0.439	0.500	0.441	0.361	0.374
A-Kn	X	**0.889**	**0.483**	0.600	**0.535**	**0.472**	**0.476**
	Y	0.846	0.225	0.186	0.200	0.123	0.122
	Z	0.870	0.448	0.410	0.405	0.336	0.348
A-Pe	X	0.843	0.273	0.257	0.263	0.177	0.177
	Y	0.774	0.155	0.181	0.157	0.035	0.036
	Z	0.798	0.075	0.086	0.080	-0.031	-0.032
M-An	X	0.771	0.229	0.357	0.271	0.150	0.156
	Y	0.831	0.150	0.086	0.109	0.032	0.030
	Z	0.865	0.430	0.257	0.321	0.250	0.261
M-Hp	X	0.849	0.384	0.381	0.373	0.289	0.294
	Y	0.866	0.450	0.238	0.299	0.231	0.252
	Z	0.897	0.537	0.381	0.440	0.389	0.398
M-Kn	X	<u>**0.962**</u>	<u>**0.893**</u>	<u>**0.800**</u>	<u>**0.819**</u>	<u>**0.799**</u>	<u>**0.814**</u>
	Y	0.894	0.587	0.324	0.414	0.361	0.382
	Z	0.883	0.540	0.238	0.301	0.251	0.287
Ensemble	HV	0.895	0.400	0.086	0.139	0.128	0.173
	SV	0.895	0.350	0.114	0.159	0.146	0.177
	ST	**0.908**	**0.760**	**0.233**	**0.329**	**0.307**	**0.380**

Regarding the proposed ensemble architecture, the results show that it failed to outperform its base classifiers regarding detection of the gait pattern anomalies. Moreover, more complex aggregation methods, such as stacking, did not consistently lead to performance improvements. Results also suggest that not all features are equally useful in distinguishing between normal and anomalous gait patterns, as suggested by the base classifiers performance. Furthermore, the usefulness of the various features varies by problem, meaning that specific gait anomallies may be better described by different sets of features.

Attending to the general adequacy of similar approaches that was found in the literature for similarly structured problems, we conclude the under-performance described in this work is related to the challenges posed by the dataset used. Concretely, we propose that the limited number of samples, high class imbalance, and variable feature importance may be the main contributors degrading the ensemble performance. In order to better understand the experimental results, a more detailed study of the ensemble performance would be required, ideally comparing experimental results for various multi-view datasets with distinct features. Furthermore, it would also be relevant to study the performance impact of different methods for base model identification, including other fuzzy approaches.

Acknowledgments. This work was supported by Fundação para a Ciência e Tecnologia (FCT) under the PhD scholarship 2022.14216.BDANA, and through IDMEC, under LAETA, to whom the authors express their gratitude. Moreover, this work was also made possible by Faculdade de Motricidade Humana (FMH) and Prof. Filipa João[0000−0002−8976−0932], to whom the authors express their sincerest gratitude for providing access to the used dataset.

References

1. L. A. Rosenbaum P, Paneth N, A report: the definition and classification of cerebral palsy, Developmental Medicine & Child Neurology 49 (s109) (2007) 8–14. ,
2. J. M. Rodda, H. K. Graham, L. Carson, M. P. Galea, R. Wolfe, Sagittal gait patterns in spastic diplegia, The Journal of Bone and Joint Surgery 86-B-British Volume (2) (2004) 251–258.
3. A. Leardini, C. Belvedere, F. Nardini, N. Sancisi, M. Conconi, V. Parenti-Castelli, Kinematic models of lower limb joints for musculo-skeletal modelling and optimization in gait analysis, Journal of Biomechanics 62 (2017) 77–86, human Movement Analysis: The Soft Tissue Artefact Issue.
4. F. Joao, V. Bagao, S. Cabral, A. Veloso, P 153 - variability of gait parameters in children: The importance of normative data establishment, Gait & Posture 65 (2018) 488–489, eSMAC 2018 abstracts.
5. V. Moniz-Pereira, T. M. Kepple, S. Cabral, F. João, A. P. Veloso, Joint moments' contributions to vertically accelerate the center of mass during stair ambulation in the elderly: An induced acceleration approach, Journal of Biomechanics 79 (2018) 105–111.
6. M. Karatas, L. Eriskin, M. Deveci, D. Pamucar, H. Garg, Big data for healthcare industry 4.0: Applications, challenges and future perspectives, Expert Systems with Applications 200 (2022) 116912.
7. Q. Tang, J. Liang, F. Zhu, A comparative review on multi-modal sensors fusion based on deep learning, Signal Processing 213 (2023) 109165.
8. S. Shi, F. Nie, R. Wang, X. Li, When multi-view classification meets ensemble learning, Neurocomputing 490 (2022) 17–29.
9. R. Marcinkevics, J. E. Vogt, Interpretability and explainability: A machine learning zoo mini-tour, ArXiv abs/2012.01805 (2020).
10. P. Angelov, X. Gu, Autonomous learning multi-model classifier of 0-order (almmo-0), in: 2017 Evolving and Adaptive Intelligent Systems (EAIS), 2017, pp. 1–7.
11. E. Soares, P. Angelov, X. Gu, Autonomous learning multiple-model zero-order classifier for heart sound classification, Applied Soft Computing 94 (2020) 106449.
12. P. Angelov, X. Gu, Mice: Multi-layer multi-model images classifier ensemble, in: 2017 3rd IEEE International Conference on Cybernetics (CYBCONF), 2017, pp. 1–8.
13. P. Angelov, X. Gu, D. Kangin, J. Principe, Empirical data analysis: A new tool for data analytics, in: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2016, pp. 000052–000059.

14. P. Angelov, R. Yager, A new type of simplified fuzzy rule-based system, International Journal of General Systems 41 (2) (2012) 163–185. ,
15. T. Takagi, M. Sugeno, Fuzzy identification of systems and its applications to modeling and control, IEEE Transactions on Systems, Man, and Cybernetics SMC-15 (1) (1985) 116–132.
16. E. H. Mamdani, Applications of fuzzy algorithms for control of a simple dynamic plant, Proceedings of the IEEE (1974).