applied sciences

MDPI

Article

Post-Stroke Gait Classification Based on Feature Space Transformation and Data Labeling

Robert Burduk [1,*], Izabela Rojek [2], Emilia Mikołajewska [3] and Dariusz Mikołajewski [2]

[1] Wroclaw University of Science and Technology, Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland
[2] Kazimierz Wielki University, Chodkiewicza 30, 85-064 Bydgoszcz, Poland
[3] Nicolaus Copernicus University Collegium Medicum in Bydgoszcz, Jagiellońska 13-15, 85-067 Bydgoszcz, Poland
* Correspondence: robert.burduk@pwr.edu.pl

Abstract: Despite scientific and clinical advances, stroke is still considered one of the main causes of disability, including gait disorders. The search for more effective methods of gait re-education in post-stroke patients is one of the most important issues in contemporary neurorehabilitation. In this paper, we propose a transformation of the feature space and definition of class labels in the post-stroke gait problem to more efficiently study related phenomena and assess gait faster. Clustering is used to define two class labels (improvement and recurrence) in the data labeling process. The proposed approach was tested on a real-world dataset consisting of 50 patients (male and female, aged 49–82 years) after ischemic stroke who participated in a gait rehabilitation program. Gait in the study was described using speed, cadence, and stride length and their normalized values. Ten treatment sessions (10 therapy days) were conducted over two weeks (10 working days). The same specialist took measurements, and hence inter-rater reliability can be neglected. Machine learning methods, support vector machine and quadratic discriminant analysis were used to classify post-stroke gait for three cases with different class labels. The proposed novel approach, characterized by its speed of execution and accuracy of classification, may be helpful for screening, better targeting, and rehabilitation monitoring. The proposed approach minimizes clinical testing and supports the work of physicians, physiotherapists, and diagnosticians.

Keywords: clinical gait analysis; stroke; rehabilitation; machine learning

Citation: Burduk, R.; Rojek, I.; Mikołajewska, E.; Mikołajewski, D. Post-Stroke Gait Classification Based on Feature Space Transformation and Data Labeling. *Appl. Sci.* **2022**, *12*, 11346. https://doi.org/10.3390/app122211346

Academic Editor: Hanatsu Nagano

Received: 15 October 2022
Accepted: 4 November 2022
Published: 8 November 2022

1. Introduction

Stroke incidence and prevalence can be considered stabilized, especially in developed countries, but stroke continues to be one of the three most common medical conditions in the world. This situation calls for increased stroke surveillance and care, but above all, a better management of risk factors, mainly undetected or uncontrolled hypertension (HBP). Estimates of stroke incidence among stroke survivors vary considerably in low-income countries in both magnitude and trend. Health-related characteristics (receipt of health services) were the main correlates of stroke in older women, whereas lifestyle and post-stroke behaviors are the most significant correlates for younger stroke survivors [1,2].

The sequelae of stroke, including mild to severe deficits and disability as well as loss of independence, significantly affect the quality of life of stroke patients and their families/caregivers [3]. Walking dysfunctions in part persist for many months after rehabilitation following stroke. A major limitation of current rehabilitation approaches is the inability to identify modifiable deficits whose improvement will result in a return of gait function. For the aforementioned reasons, researchers are turning to computational neuroscience to accelerate and facilitate our understanding of the mechanisms involved [4–6].

Our previous studies showed that even after the short-term (2-week) post-stroke rehabilitation, benefits were observed in terms of measurable, statistically significant

changes in the patients' gait parameters, changes in hand function, and in activities of daily living; however, it is difficult to achieve a simultaneous improvement in all areas: gait parameters, hand function, and ADLs within two weeks of rehabilitation [7–9].

The advent of many new therapeutic approaches as well as the effects of the COVID-19 pandemic have brought with them gaps in contemporary gait analysis associated with the need for prolonged direct contact, hence the need for new, more automated approaches, including rapid, cost-effective, accurate, and useful methods for collecting second opinions in clinical gait analysis.

A review of three major bibliographic databases (PubMed, PEDro, Scopus) using specified keywords (gait analysis, machine learning (ML), physiotherapy, among others) showed that between 2009 and 2022, only 20 papers were published on ML applications in clinical gait analysis. Only 2 of the papers were related to gait analysis [10,11]. Both indicated that the use of ML tools in clinical gait analysis in post-stroke patients is insufficient and requires further research. This indicates an important and identified research and clinical gap, which our study is now filling.

The main novelties and contributions of the presented methods are the speed of execution, the reproducibility and accuracy of the classification, and a better fit with daily clinical practice than a traditional gait laboratory. Additionally, these will support doctors and physiotherapists in their daily work with patients without needing sophisticated and expensive hardware and software. Ultimately, this type of software can be installed on mobile devices (tablets, smartphones) and for home rehabilitation or telerehabilitation.

This work aimed to propose feature space transformation that allows the definition of new class labels in the post-stroke classification problem. The new feature space defines two class labels, improvement and relapse, based on clustering. The proposed approach serves to minimize clinical testing, as well as support the work of physicians, physiotherapists, and diagnosticians. Machine learning methods, support vector machine (SVM) and quadratic discriminant analysis (QDA) were used for post-stroke gait classification for three cases with different class labels. One of these cases is related to previously defined new class labels.

2. Material and Methods

2.1. Material

This study group consisted of 50 patients after ischemic stroke who participated in the rehabilitation program conducted by the Department of Physiotherapy at Collegium Medicum im. Ludwik Rydygiera in Bydgoszcz, Nicolaus Copernicus University in Toruń and the company Fizjoterapia Emilia Mikołajewska in 2020–2021. The method of patient recruitment: convenience sample. The research groups were set up on described criteria using a traditional patient flow diagram and draw for groups. The inclusion criteria were as follows: age over 18 years; medical status each time confirmed by medical documentation; post-stroke time over 1 month to 3 years; lack of contraindications to rehabilitation procedures. The appropriate bioethical commission approved this study, giving written informed consent prior to testing, operating according to the principles of Good Clinical Practice and the Helsinki Declaration.

The outcomes analyzed were:

- In the study group (n = 25): gait rehabilitation results of adult patients after ischemic stroke (gait parameters before and after rehabilitation);
- In the reference group (n = 25): gait parameters in healthy subjects.

Data of both groups are presented in Table 1.

Table 1. Clinical summary of the patients.

	Study Group (n = 25, 100%)	Reference Group (n = 25, 100%)
	Age (years)	
Mean	65.28	58.6
SD	9.65	6.29
Min	49	51
Q1	57	54
Median	68	56
Q3	74	61
Max	82	72
	Side of paresis:	
Left (L)	13 (52%)	n. a.
Right (R)	12 (48%)	n. a.
	Sex:	
Female (F)	12 (48%)	13 (52%)
Male (M)	13 (52%)	12 (48%)

n.a.—not applicable.

Where:

- Q1—lower quartile, or first quartile (Q1), is the value under which 25% of data points are found when they are arranged in increasing order;
- Q2—upper quartile, or third quartile (Q3), is the value under which 75% of data points are found when arranged in increasing order;
- SD—standard deviation.

2.2. Methods

Gait evaluation was based on spatial-temporal gait parameters (gait speed, cadence, and stride length), i.e., their normalized parameters—related to lower limb length, which allows their comparison between patients of different height and stride length. In the healthy group, parameter measurements were performed by the same therapist once during the 10 m walking test. In the sick group, the parameters were measured by the same therapist twice during the 10 m walking test: on admission to the hospital and after the tenth session of gait re-education in order to evaluate the effects of rehabilitation. The study group consisted of 25 patients after ischemic stroke who participated in a rehabilitation program. Ten therapy sessions (10 days of therapy) were conducted over a 2-week period (ten working days). Measurements were made by the same specialist, and hence inter-rater reliability may be omitted.

To summarize, each patient was subjected to a 10 m walk test (10 MWT), which was recorded using a digital smartphone camera. The ground contact time etc., was calculated and used to calculate the velocity, cadence, and stride length in relation to their normalized values using Chris Kirltey's Clinical Gait Analyzer software, which is available online.

To obtain the leg length, the examiner measures from the anterior superior iliac spine (ASIS) to the medial malleolus. By entering the leg length (ASIS to medial malleolus), you can see how close the patient is to normal and allow you to compare gait parameters in patients with various leg lengths. Thus, velocity, cadence, and stride length after the normalization procedure using leg length become normalized velocity, normalized cadence, and normalized stride length. Normalization was performed using Clinical Gait Analysis software by Chris Kirtley (http://www.clinicalgaitanalysis.com/(accessed on 1 June 2022)).

Hardware and software tools employed in our work:

- Clinical Gait Analysis software by Chris Kirtley (available online at http://www.clinicalgaitanalysis.com/(accessed on 1 June 2022)) for gait parameters calculation purposes;
- Statistica 13 (StatSoft, Tulsa, OK, USA) for statistical analysis purposes;

- SAS 9.4 (SAS Institute, Cary, NC, USA) software was used to conduct computational analyses.

The results were subjected to statistical and computational analyses. All the data were analyzed with Statistica version 13 software. The parameter change was defined as the result of subtraction. The distribution of values was determined using the Shapiro–Wilk test. Values for distributions close to the normal distribution were presented by mean values and standard deviation (SD). Values for distributions different from the normal distribution were presented by median, minimum value, maximum value, lower quartile, and upper quartile. The T-student test was used to compare the results. Spearmans rho was used to extract correlations. The significance level was set at 0.05.

2.3. Nomenclature Used in This Article

The following nomenclature is used in the article: Clinical gait parameters:

- Velocity—The speed of gait in a given direction;
- Cadence—Total number of full cycles taken within a given period of time, i.e., here the number of steps taken per minute;
- Stride length—The distance between successive points of initial contact of the same foot, i.e., distance covered when you take two steps, one with each foot.

and their normalized (i.e., taking into consideration leg length from ASIS to medial malleolus) values, i.e.:

- normalized velocity;
- normalized cadence;
- normalized stride length.

3. Therapy Outcomes

The results of the study and reference group are presented in Table 2. After the short-term therapy, the benefits expressed in measurable, statistically significant changes in patients' gait parameters were observed and presented in Tables 3 and 4.

Among the patients participating in the study, the following results were obtained: in terms of gait speed, recovery was observed in 52% cases; in terms of cadence, recovery was observed in 48% cases; and in terms of stride length, recovery was observed in 76% of cases.

Table 2. Results for the study and reference group.

	Normalized Velocity	Normalized Cadence	Normalized Stride Length
Before therapy—study group			
Mean	0.16	0.39	1.62
SD	0.07	0.09	0.57
Min	0.05	0.17	0.38
Q1	0.12	0.34	1.34
Median	0.16	0.41	1.73
Q3	0.2	0.45	1.98
Max	0.28	0.51	2.5
After therapy—study group			
Mean	1.19	0.41	1.86
SD	0.12	0.16	0.62
Min	0.04	0.12	0.72
Q1	0.12	0.27	1.57
Median	0.17	0.43	1.85
Q3	0.25	0.51	2.17
Max	0.53	0.76	2.94

Table 2. *Cont.*

	Normalized Velocity	Normalized Cadence	Normalized Stride Length
	Reference group		
Mean	0.606	0.618	1.931
SD	0.061	0.054	1.211
Min	0.52	0.54	1.7
Q1	0.56	0.57	1.745
Median	0.58	0.61	1.835
Q3	0.66	0.66	2.177
Max	0.72	0.73	2.23

Table 3. The general effect of rehabilitation in the study group (n = 25).

	Normalized Velocity	Normalized Cadence	Normalized Stride Length
Recovery	13 (52%)	12 (48%)	19 (76%)
No change	3 (12%)	2 (8%)	2 (8%)
Relapse	9 (36%)	11 (44%)	4 (16%)

Table 4. The change in parameters' values reflecting the effect of rehabilitation in the study group.

	Normalized Velocity	Normalized Cadence	Normalized Stride Length
Mean	0.0364	0.02	0.24
SD	0.008	0.006	0.072
Min	−0.07	−0.18	−0.6
Q1	−0.02	−0.06	0.11
Median	0.01	0	0.19
Q3	0.08	0.09	0.38
Max	0.25	0.34	1.19
p-value	0.047	0.033	0.022

4. Feature Space Transformation and Data Labeling

The patient (instance) before therapy is represented in feature space as the three-dimensional feature vector $x = [nv(bt), nc(bt), nsl(bt)]$ where $nv(bt)$—normalized velocity; $nc(bt)$—normalized cadence; and $nsl(bt)$—normalized stride length before therapy, respectively. After the second patient investigation (after therapy), the feature space will be increased by another three dimensions which are $nv(at)$—normalized velocity; $nc(at)$—normalized cadence; and $nsl(at)$—normalized stride length after therapy, respectively.

We propose to define three subspaces of feature space related to different types of measurements (normalized velocity, normalized cadence, and normalized stride length). The subspaces are as follows: $[nv(bt), nv(at)]$, $[nc(bt), nc(at)]$, $[nls(bt), nls(at)]$. The domain knowledge about the treatment might suggest that the difference between measurement values before and after therapy is important, and hence we propose to define a new feature subspace by mapping according to the formula:

$$\Phi[fs(bt), fs(at)] \rightarrow [fs(at), fs(at) - fs(bt)]', \quad (1)$$

where $fs \in \{nv, nc, nls\}$.

To define new domain knowledge concepts (class labels), we propose the use of the clustering method [12]. The cluster centroids can be taken as the representation of the domain knowledge concepts in the feature subspace. Those clusters centroids are abstract instances, while the actual instances closest to the centroid of a given cluster have a label like this centroid. The class labels of the centroids are manually determined by an expert. The procedure of the proposed features subspaces transformation and class labels definition is presented in Algorithm 1.

Algorithm 1: Algorithm of cases labeling.

Input : The number of the class labels – the clusters number K, set of feature subspaces fs

Output: Labeled cases

foreach $fs \in \{nv, nc, nls\}$

do

  Mapping the feature subspace $\Phi[fs(bt), fs(at)] \rightarrow [fs(at), fs(at) - fs(bt)]'$

  Cluster the feature subspace fs' by K-means algorithm

  Output: Label cases by clusters in fs'

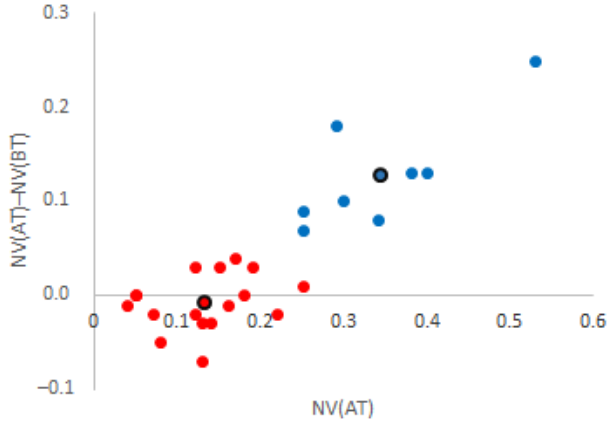
end

The labeling results for the two new concepts are shown in Figure 1 for each of the three feature subspaces. In this case, we defined two new concepts (class labels): *improvement* and *relapse* (abbreviation im and re in Table 5). The first class label indicates the patient's improvement after treatment, while the second class label indicates no improvement in the patient's condition. In Figure 1, the *improvement* class label is assigned to the blue points while the red points indicate the objects with the *relapse* class label in each feature subspace. The points with a black outline represent centroids obtained by the K-means algorithm.

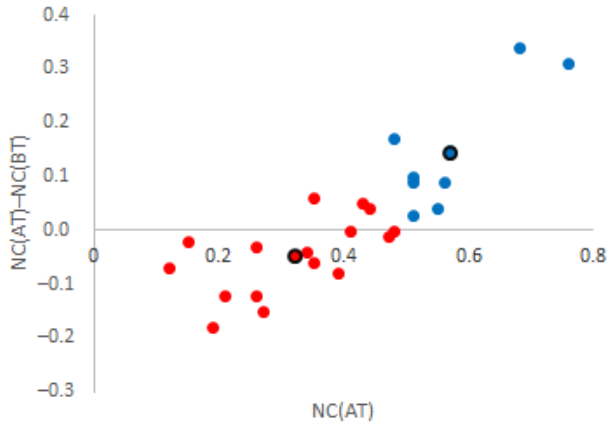
Table 5. Original features, created features, and created class labels.

$nv(bt)$	$nv(at)$	$nv(at) - nv(bt)$	$nc(bt)$	$nc(at)$	$nc(at) - nc(bt)$	$nls(bt)$	$nls(at)$	$nls(at) - nls(bt)$	Label in nv	Label in nc	Label in nls
0.11	0.29	0.18	0.31	0.48	0.17	1.42	2.41	0.99	im	im	im
0.2	0.3	0.1	0.34	0.68	0.34	2.39	1.79	−0.6	im	im	re
0.05	0.05	0	0.29	0.26	−0.03	0.72	0.72	0	re	re	re
0.16	0.25	0.09	0.42	0.57	0.15	1.55	1.79	0.24	im	im	re
0.18	0.18	0	0.41	0.35	−0.06	1.79	2.11	0.32	re	re	im
0.16	0.19	0.03	0.41	0.41	0	1.59	1.85	0.26	re	re	re
0.05	0.04	−0.01	0.33	0.21	−0.12	0.66	0.79	0.13	re	re	re
0.12	0.15	0.03	0.29	0.35	0.06	1.63	1.74	0.11	re	re	re
0.17	0.14	−0.03	0.38	0.34	−0.04	1.85	1.71	−0.14	re	re	re
0.26	0.34	0.08	0.48	0.51	0.03	2.17	2.72	0.55	im	im	im
0.14	0.12	−0.02	0.4	0.44	0.04	1.42	1.14	−0.28	re	re	re
0.05	0.05	0	0.17	0.15	−0.02	1.11	1.28	0.17	re	re	re
0.09	0.12	0.03	0.51	0.55	0.04	0.72	0.85	0.13	**re**	**im**	re
0.09	0.07	−0.02	0.19	0.12	−0.07	1.92	2.3	0.38	re	re	im
0.28	0.53	0.25	0.45	0.76	0.31	2.5	2.81	0.31	im	im	im
0.24	0.22	−0.02	0.48	0.48	0	2.03	1.88	−0.15	re	re	re
0.2	0.13	−0.07	0.42	0.27	−0.15	1.9	1.9	0	re	re	re
0.16	0.13	−0.03	0.38	0.26	−0.12	1.73	2.04	0.31	re	re	im
0.13	0.08	−0.05	0.37	0.19	−0.18	0.38	1.57	1.19	re	re	re
0.24	0.25	0.01	0.48	0.47	−0.01	1.98	2.17	0.19	re	re	im
0.13	0.17	0.04	0.38	0.43	0.05	1.21	1.63	0.42	re	re	re
0.27	0.4	0.13	0.47	0.56	0.09	2.3	2.87	0.57	im	im	im
0.17	0.16	−0.01	0.47	0.39	−0.08	1.34	1.52	0.18	re	re	re
0.18	0.25	0.07	0.41	0.51	0.1	1.79	1.94	0.15	im	im	re
0.25	0.38	0.13	0.42	0.51	0.09	2.35	2.94	0.59	im	im	im

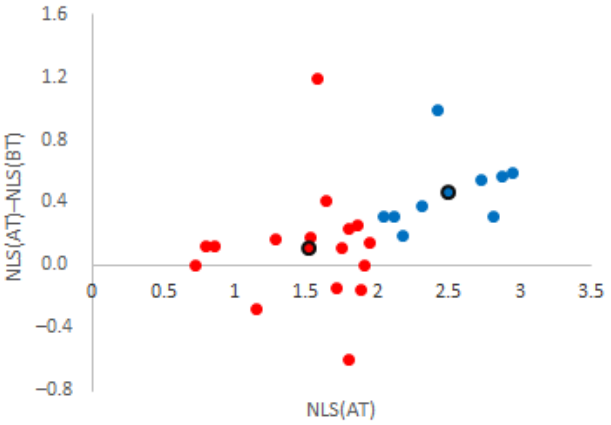
The obtained results show that instances are just as labeled in normalized velocity and normalized cadence feature subspace. Only one of the 25 instances has an assigned different label in normalized velocity and normalized cadence feature subspaces—marked in bold in Table 5. Thus, one of these two features can be discarded for practical reasons (e.g., simpler and faster feature acquisition). The conclusion about choosing one feature instead of two (normalized velocity and normalized cadence) is very interesting—the proposed approach to features' and class labels' definition may contribute to the feature selection necessary to determine whether the therapy would be effective.



(**a**) Normalized velocity subspace



(**b**) Normalized cadence subspace



(**c**) Normalized stride length subspace

Figure 1. The result of clustering for 2 clusters. The points with a black outline represent cluster centroids.

The differences in the resulting class labels between the first, second, and third feature subspaces are apparent. For this reason, in our opinion, in the classification process with new class labels, each subspace should be considered separately.

5. Supervised Classification

In the supervised classification, the dataset should contain an attribute that identifies the class label. In the case of the analyzed patient dataset, we can define three supervised classification tasks. For the first problem, the class labels are: *sick* and *healthy*. Physicians

defined these class labels in assessing the conditions of patients. In the second problem, the class labels are: *before therapy* and *after therapy*. These class labels were determined according to the measurement features of sick patients before and after therapy. In the third case, the class labels were determined by the algorithm 1 proposed in this paper. These are the class labels *improvement* and *relapse*.

where:

- improvement—when the result of the measure after therapy was higher than that before therapy;
- relapse—when the result of the measure after therapy was lower than that before therapy.

5.1. Experiment Setup

We used the SVM and QDA algorithms as a classification model. The SVM provided a range of kernels and hyperparameters for the selection [13]. In the experiments, we took into account the linear, radial basis function, polynomial and sigmoid kernels. The regularization parameter C was searched in the set $C \in \{0.001, 0.01, 0.1, 1, 10, 100\}$, and the kernel γ parameter was searched in the set $\gamma \in \{1, 5, 10, 15, 20\}$. The grid search was used to determine which kernel and hyperparameters are "appropriate" for the given supervised classification task. For example, the classifier marked in Table 6 as $\Psi_{Pol}^{1,5}$ is the SVM model with the polynomial kernel, $C = 1$ and $\gamma = 5$.

The implementation of SVM and QDA from SAS 9.4 software was used to conduct the experiments. Since the dataset is small, the number of observations n = 50, and we use the leave-one-out cross-validation method to model evaluation. A performance classification metric such as the area under the curve (AUC), the true positive rate (TPR—sensitivity), and the true negative rate (TNR—specificity) were used.

5.2. Sick vs. Healthy Classification

In this experiment, we consider patients' reference group labeled as *healthy* and patients within the study group labeled as *sick*. Each group consists of 25 patients. The statistics for the features are presented in Tables 2–4. From a supervised classification point of view, this is an easy dataset as most algorithms achieve the ideal classification performance. For example, the QDA algorithm Ψ_{QDA} has a classification performance metric AUC = 1. This proves that they consider the investigated features: normalized velocity, normalized cadence, and normalized stride length, which we can distinguish between sick and healthy people with 100% efficiency.

5.3. Before Therapy vs. after Therapy Classification

In this experiment, we consider patients previously labeled as *sick*. Patients in this group were investigated twice. Therefore, we distinguish two class labels *before therapy* and *after therapy*. This classification problem is much more complicated and it was required to find an appropriate hyperparameter set for the SVM classifier. The results of the experiments are presented in Table 6. The TPR is related to the *before therapy* class label and the FPR is related to the *after therapy* class label.

Table 6. Result of classification—before therapy vs. after therapy.

	Classification Algorithm				
Metric	$\Psi_{L}^{0.01}$	$\Psi_{RBF}^{0.01,15}$	$\Psi_{Pol}^{1,5}$	$\Psi_{Sig}^{1,5}$	Ψ_{QDA}
AUC	0.76	0.9	0.58	0.5	0.6
TPR	0.88	1	0.84	1	0.72
FPR	0.64	0.8	0.32	0	0.48

The obtained results confirm that the hyperparameter and the kernel function selection is essential to the robust classification performance of the SVM algorithm. For the analyzed supervised classification problem, the $\Psi_{RBF}^{0.01,15}$ algorithm with the radial basis function

kernel, $C = 0.01$ and $\gamma = 15$ is the best one in each of the considered performance classification metrics. In particular, TPR=1, which means that with 100% efficiency, it is possible to recognize patients before therapy with a reasonably high error coefficient expressed by AUC = 0.9.

In the conducted research, we also analyzed the validity of using principal component analysis for the dimensionality reduction. Note that the application of this method did not improve the classification results.

5.4. Improvement vs. Relapse Classification

In this supervised classification problem, most algorithms achieve the ideal classification performance (see Table 7). It is evident for the normalized velocity feature subspace, where algorithm Ψ_{QDA} also obtained the highest value of the classification performance metrics. In the case of normalized stride length feature subspace, the results are slightly worse. However, it should be noted that the value of the TPR metric equals 1. Therefore, in this feature subspace, the objects belonging to the *improvement* class label are correctly classified.

Table 7. Result of classification—improvement vs. relapse.

	Classification Algorithm				
	Ψ_{L}^{100}	$\Psi_{RBF}^{100,1}$	$\Psi_{Pol}^{100,10}$	$\Psi_{Sig}^{1,5}$	Ψ_{QDA}
Metric	Normalized velocity subspace				
AUC	1	1	1	0.5	1
TPR	1	1	1	1	1
FPR	1	1	1	0	1
Metric	Normalized cadence subspace				
AUC	1	1	1	0.5	0.9
TPR	1	1	1	1	1
FPR	1	1	1	0	0.81
Metric	Normalized stride length subspace				
AUC	0.96	0.96	0.96	0.53	0.94
TPR	1	1	1	1	1
FPR	0.93	0.93	0.93	0.06	0.88

6. Discussion

Physiological and pathological human gait is a very complex and fluid set of movements, requiring the synchronous activity of several body systems, difficult to describe unambiguously and evaluate objectively. New, more reliable classifiers for computational gait analysis after stroke were presented. We view our results as preliminary. Although promising, further studies on larger samples are needed.

6.1. Compartment to the Results from the Other Studies

The ongoing pandemic poses new challenges to gait diagnosis and therapy. Limitations in access to medical specialists and gait laboratories and the prospect of the abovementioned situation being indefinitely prolonged into the future prioritizes fast, low-cost and telemedically and telerehabilitation-enabled clinical gait analysis procedures. As strokes are not the only conditions resulting in gait function deficits, the aforementioned tools should investigate gait deficits resulting from other conditions with similar efficiency levels, such as neurodegenerative changes in the elderly, effects of craniocerebral and spinal cord injuries, severe poisoning, metabolic diseases, and burns. They should also be suitable for use by general practitioners, physiotherapists and, in special situations, by patients themselves at home thanks to the automation of the measurement process in the framework of, for example, the 10 m walking test and a quick dichotomous result: a healthy patient or a patient with a deficit (to be further diagnosed in a more complex manner). This would relieve the burden on the healthcare system without compromising the quality of

gait diagnosis. In our opinion, the solution we demonstrated meets the requirements of being a faster solution than the artificial neural networks, fuzzy analysis, or fractal analysis used to date, and it is therefore possible to use on mobile devices (52% of the Internet access in the world occurs through mobile devices). A comparison with the results of other researchers is not straightforward, as practical, rapid, and low-cost studies are scarce [5–14]. The existing indices of gait functionality, including post-stroke, can describe the distance of gait from normal and help to monitor the return of gait to normal as a result of self-healing or recovery processes in the body supported by the rehabilitation processes by observing improvements in gait parameters (e.g., stride length, durations of gait cycle components, etc.). However, these indices reveal little information related to gait quality or kinesiology status [15], and are not quick and cheap to use as screening tests. This overlaps with the need to more fully understand the process of human motor control by developing a model of behavior in different situations (including gait disturbed by external factors) in patients with postural disorders [16].

6.2. Limitations of Own Studies

The limitations of our research stem from its assumed nature and include: a homogeneous group of post-stroke patients, a small sample and the way it was selected (a convenience sample). The computational nature of the study and the way the classifiers were selected seems to be effective—for this purpose, we made a comparison with an independent gait diagnostician who evaluated the patients' performance based on their own experience, and our classification system met the aforementioned evaluation, achieving results comparable to those of a human specialist. A specialist's judgement is not only the product of their knowledge and experience, but also that of clinical intuition. It is a kind of Turing test for clinical software.

6.3. Directions for Further Research

We treat our research as a preliminary study. Directions for further research primarily include:

- The evaluation of larger groups of patients, including walking and sitting activities [17,18];
- The evaluation of groups of patients with gait deficits caused by other diseases and injuries (e.g., neuromotor disabilities in children [19], hand deficits [20], and robotic therapy [21,22])—on this basis, the expansion of the classifier base to classifiers that optimally reflect the specificity of individual patient groups;
- The development of a fully functional system that can be used by non-specialists in clinical settings.

7. Conclusions

In this study, the following research steps were completed: a literature review was conducted to establish the state of the art, a research gap was found, data from original studies were extracted, specific computational analyses were performed using selected ML methods, and the author's approach allowed for better results leading to the automation and support for rehabilitation therapists. The proposed feature space transformation and data labeling method minimizes clinical testing and supports the work of doctors, physiotherapists, and diagnosticians in post-stroke gait problems. The proposed approach allows the earlier and less error-prone detection of patients with gait deficits of different causes, earlier implementation of appropriate therapy, and better targeting and monitoring of rehabilitation. Ultimately, they will contribute to this patient's increased quality of life in this group of patients.

Author Contributions: Conceptualization, I.R., R.B., D.M. and E.M.; methodology, I.R., R.B., D.M. and E.M.; software, R.B.; validation, I.R., R.B. and D.M. formal analysis, I.R., R.B., D.M. and E.M.; investigation, I.R., R.B. and D.M. resources, I.R., D.M. and E.M.; data curation, I.R., R.B., D.M. and E.M.; writing—original draft preparation, I.R., R.B., D.M. and E.M.; writing—review and editing, I.R., R.B., D.M. and E.M.; visualization, I.R., R.B. and D.M.; supervision, I.R., R.B. and E.M.; project administration, I.R. and R.B.; funding acquisition, I.R., R.B., D.M. and E.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the grant to maintain the research potential of Kazimierz Wielki University (Ministry of Education and Science, grant 2022).

Institutional Review Board Statement: Consent of the Bioethics Committee of Ludwik Rydygier Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Toruń, Poland no. KB 634/2017.

Informed Consent Statement: All study participants gave a Statement of Informed Consent.

Data Availability Statement: The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ezejimofor, M.C.; Chen, Y.F.; Kandala, N.B.; Ezejimofor, B.C.; Ezeabasili, A.C.; Stranges, S.; Uthman, O.A. Stroke survivors in low-and middle-income countries: A meta-analysis of prevalence and secular trends. *J. Neurol. Sci.* **2016**, *364*, 68–76. [PubMed]
2. Jacobs, M.M.; Ellis, C. Heterogeneity among women with stroke: Health, demographic and healthcare utilization differentials. *Bmc Women'S Health* **2021**, *21*, 160. [CrossRef] [PubMed]
3. Aguilera, A.; Subero, A. Automatic gait classification patterns in spastic hemiplegia. *Adv. Data Anal. Classif.* **2020**, *14*, 897–925. [CrossRef]
4. Appukutty, A.J.; Skolarus, L.E.; Springer, M.V.; Meurer, W.J.; Burke, J.F. Increasing false positive diagnoses may lead to overestimation of stroke incidence, particularly in the young: A cross-sectional study. *BMC Neurol.* **2021**, *21*, 152. [CrossRef] [PubMed]
5. Hollands, K.L.; Pelton, T.A.; Tyson, S.F.; Hollands, M.A.; van Vliet, P.M. Interventions for coordination of walking following stroke: Systematic review. *Gait Posture* **2012**, *35*, 349–359. [CrossRef]
6. Richards, C.L.; Malouin, F.; Dean, C. Gait in stroke: Assessment and rehabilitation. *Clin. Geriatr. Med.* **1999**, *15*, 833–856.
7. Prokopowicz, P.; Mikołajewski, D.; Tyburek, K.; Mikołajewska, E. Computational gait analysis for post-stroke rehabilitation purposes using fuzzy numbers, fractal dimension and neural networks. *Bull. Pol. Acad. Sci. Tech. Sci.* **2020**, *68* , 191–198.
8. Mikołajewska, E. Associations between Results of Post-Stroke NDT-Bobath Rehabilitation in Gait Parameters, ADL and Hand Functions. *Adv. Clin. Exp. Med.* **2013**, *22*, 731–738.
9. Mikołajewska, E. The value of the NDT-Bobath method in post-stroke gait training. *Adv. Clin. Exp. Med.* **2013**, *22*, 261–272.
10. Mohan, D.; Khandoker, A.; Wasti, S.; Ibrahim, I.; Alali, S.; Jelinek, H.F.; Khalaf, K. Assessment Methods of Post-stroke Gait: A Scoping Review of Technology-Driven Approaches to Gait Characterization and Analysis. *Front. Neurol.* **2021**, *12*, 650024. [CrossRef]
11. Luvizutto, G.; Silva, G.; Nascimento, M.; Sousa Santos, K.; Appelt, P.; de Moura Neto, E.; de Souza, J.; Wincker, F.; Miranda, L.A.; Hamamoto Filho, P.; et al. Use of artificial intelligence as an instrument of evaluation after stroke: A scoping review based on international classification of functioning, disability and health concept. *Top Stroke Rehabil.* **2022**, *29*, 331–346. [PubMed]
12. Homenda, W.; Jastrzebska, A. Clustering techniques for Fuzzy Cognitive Map design for time series modeling. *Neurocomputing* **2017**, *232*, 3–15. [CrossRef]
13. Scholkopf, B.; Smola, A.J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; Adaptive Computation and Machine Learning Series; MIT Press: Cambridge, MA, USA, 2018.
14. Wonsetler, E.C.; Bowden, M.G. A systematic review of mechanisms of gait speed change post-stroke. Part 2: Exercise capacity, muscle activation, kinetics, and kinematics. *Top. Stroke Rehabil.* **2017**, *24*, 394–403. [CrossRef] [PubMed]
15. Tan, M.G.; Ho, J.H.; Goh, H.T.; Ng, H.K.; Latif, L.A.; Mazlan, M. Revealing stroke survivor gait deficits during rehabilitation using ensemble empirical mode decomposition of surface electromyography signals. *Biomed. Signal Process. Control.* **2020**, *61*, 102045.
16. Hedjazi, N.; Benali, A.; Bouzit, M.; Dibi, Z. Model identification and evaluation of postural dynamics of healthy and post-stroke individuals under unidirectional perturbations. *Biomed. Signal Process. Control.* **2018**, *43*, 75–85. [CrossRef]
17. Yaron Katz, T.; Hallevi, H.; Molad, J.; Kafri, M. Walking and Sitting Time after a Stroke: A Comparison of Shifts and Changes over Time within an Acute Care Setting. *Appl. Sci.* **2022**, *12*, 10945.
18. Kim, A.R.; Park, D.; Lee, Y.S. Reliability and Validity of the Ground Reaction Force Asymmetric Index at Seat-Off as a Measure of Lower Limb Functional Muscle Strength: A Preliminary Study. *Appl. Sci.* **2021**, *11*, 6527. [CrossRef]
19. Nossa, R.; Porro, M.; Picciolini, O.; Malosio, M.; Pittaccio, S.; Valoriani, M.; Asperti, V.; Clasadonte, F.; Oliveto, L.; Mondellini, M.; et al. What Children with Neuromotor Disabilities Need to Play with Technological Games. *Appl. Sci.* **2021**, *11*, 9886. [CrossRef]

20. Yang, X.; Lim, Z.; Jung, H.; Hong, Y.; Zhang, M.; Park, D.; You, H. Estimation of Finite Finger Joint Centers of Rotation Using 3D Hand Skeleton Motions Reconstructed from CT Scans. *Appl. Sci.* **2020**, *10*, 9129. [CrossRef]
21. Lin, M.; Wang, H.; Yang, C.; Liu, W.; Niu, J.; Vladareanu, L. Human–Robot Cooperative Strength Training Based on Robust Admittance Control Strategy. *Sensors* **2022**, *22*, 7746. [CrossRef]
22. Pinheiro, C.; Figueiredo, J.; Cerqueira, J.; Santos, C. Robotic Biofeedback for Post-Stroke Gait Rehabilitation: A Scoping Review. *Sensors* **2022**, *22*, 7197. [PubMed]