Article

Grammatical Evolution-Based Feature Extraction for Hemiplegia Type Detection

Vasileios Christou [1,*](ID), Ioannis Tsoulos [1], Alexandros Arjmand [1], Dimitrios Dimopoulos [2](ID), Dimitrios Varvarousis [2](ID), Alexandros T. Tzallas [1](ID), Christos Gogos [1](ID), Markos G. Tsipouras [3](ID), Evripidis Glavas [1], Avraam Ploumis [2] and Nikolaos Giannakeas [1,*](ID)

1   Department of Informatics and Telecommunications, University of Ioannina, 47100 Arta, Greece
2   Department of Physical Medicine and Rehabilitation, University of Ioannina, S. Niarchos Ave, 45110 Ioannina, Greece
3   Department of Electrical and Computer Engineering, University of Western Macedonia, 50100 Kozani, Greece
*   Correspondence: bchristou1@gmail.com (V.C.); giannakeas@uoi.gr (N.G.)

Abstract: Hemiplegia is a condition caused by brain injury and affects a significant percentage of the population. The effect of patients suffering from this condition is a varying degree of weakness, spasticity, and motor impairment to the left or right side of the body. This paper proposes an automatic feature selection and construction method based on grammatical evolution (GE) for radial basis function (RBF) networks that can classify the hemiplegia type between patients and healthy individuals. The proposed algorithm is tested in a dataset containing entries from the accelerometer sensors of the RehaGait mobile gait analysis system, which are placed in various patients' body parts. The collected data were split into 2-second windows and underwent a manual pre-processing and feature extraction stage. Then, the extracted data are presented as input to the proposed GE-based method to create new, more efficient features, which are then introduced as input to an RBF network. The paper's experimental part involved testing the proposed method with four classification methods: RBF network, multi-layer perceptron (MLP) trained with the Broyden–Fletcher–Goldfarb–Shanno (BFGS) training algorithm, support vector machine (SVM), and a GE-based parallel tool for data classification (GenClass). The test results revealed that the proposed solution had the highest classification accuracy (90.07%) compared to the other four methods.

Keywords: accelerometer; feature construction; grammatical evolution; radial basis function network

1. Introduction

Hemiplegia is caused by brain damage or spinal cord injury resulting in motor control loss in half of a person's body. The affected body area has issues, including movement inability of one arm and leg in the right or left body part area. Other recorded problems include movement in spastic patterns and stereotyped synergies. Additionally, there is selective activity loss in the muscles responsible for trunk control, specifically in the flexion, rotation, and lateral flexion muscles. Hemiplegia patients also suffer from difficulties in trunk movement regardless of the muscle action type, and the abdomen's muscles demonstrate very high activity and tone loss [1,2]. The patient's problems are not restricted only to the affected areas, they are extended to a certain degree in all brain functions since the human brain is a massive network of interconnected neurons. The person suffering from hemiplegia will face motor impairment in both body sides, affecting balance, coordination, sensory perception, and spatial orientation. The behavior, cognition, and memory will be affected, causing a significant challenge in the rehabilitation procedure [3,4].

The current article proposes an automated system for diagnosing the hemiplegia type (right or left body part is paralyzed) between healthy and non-healthy individuals. The motivation behind creating such a system is to provide an inexpensive supplementary tool that does not require expensive three-dimensional (3D) cameras for hemiplegia type

diagnosis. The proposed solution is based on the RehaGait mobile gait analysis system from HASOMED. The RehaGait mobile gait analysis system utilizes seven sensors placed in four different body parts of the individual for documenting spatio-temporal-specific data during walking. The sensors are provided by the RehaGait system, where one sensor is placed on the patient's hip and one sensor on each foot, shank, and thigh [5,6].

The selected gait analysis system includes an integrated video capture capability for observing the patient's condition. It can find the damaged areas and evaluates the gait pattern. It is suitable for mobile use, identifies asymmetries in the lower limbs, has movement freedom, and can visualize the results using a graphical representation [7,8]. Each sensor can capture signals from three different sources (accelerometer, magnetometer, and gyroscope), although the proposed system utilizes only the data taken from the accelerometer [6]. The captured data were divided into 2-second windows and underwent a pre-processing and feature extraction process. Then, they were introduced as input to the feature construction method for radial basis function (RBF) networks (FC4RBF) based on grammatical evolution (GE) by Gavrilis et al. [9]. The training procedure utilized a dataset composed of thirty individuals. The dataset is divided into ten healthy patients, eight patients with left hemiplegia, and twelve patients with right hemiplegia.

The paper contains seven sections starting with Section 1, which presents the problem's description and motivation. Section 2 documents the related work regarding existing works on hemiplegia. It also describes the algorithms used to compare the proposed method in the paper's experimental part. Section 3 describes the architecture of the selected approach for the classification task. Section 4 depicts the proposed system's analysis. Section 5 compares the method with four machine learning algorithms, while Sections 6 and 7 include the "Discussion" and "Conclusions".

2. Literature Review

Existing methods are focused on different aspects, including stroke and hemiplegic patients' gait. The authors in Patil et al. [10] proposed a machine learning method based on deep learning and convolutional neural networks (CNNs) to identify hemiplegic gait. The study Yardimci [11] used a fuzzy logic system that utilized a Tsukamoto-type inference method for the classification task between healthy and hemiplegic patients. The authors in Manca et al. [12] used hierarchical cluster analysis for classifying the gait patterns of patients having equinus foot deformity. The work in Mulroy et al. [13] used non-hierarchical cluster analysis, which utilizes the temporal/spatial and kinematic parameters of walking to categorize the gait patterns of stroke patients. The authors in Straudi et al. [14] utilized cluster analysis for identifying gait patterns from Spatio-temporal parameters, walking endurance, and sagittal plane kinematic profiles. This study aimed to categorize the walking performance of hemiplegic patients. The study Luo et al. [15] used the gait data from 20 hemiplegic and 40 healthy individuals taken from low-cost depth cameras. Then, they used the random forest (RF) classification algorithm to analyze the importance of different gait features. The authors in Zhu et al. [16] utilized the second version of Microsoft Kinetic for Windows to record the walking trajectory data from patients with hemiplegia. The data were used for gait identification feature extraction and introduced as input to a Bayesian classification algorithm for automatic recognition of hemiplegic gait. Finally, the RF algorithm was selected to identify each feature's significance. Pauk and Minta-Bielecka [17] conducted a gait analysis by comparing clustering and bi-clustering algorithms in 41 patients with hemiplegia. Moreover, they proposed a new bi-clustering algorithm named KMB. Wang et al. [18] recorded body surface electrical signals from 16 stroke patients having various degrees of rehabilitation and eight healthy individuals. These signals were filtered and underwent a wavelet transform procedure for extracting features. Then, they were introduced into linear regression (LR), ridge regression (RR), extreme gradient boosting (XGBoost), and support vector machine (SVM) models for condition evaluation and classification (healthy, mild, or severe hemiplegia). The work in Wong et al. [19] made a foot contact pattern analysis in 60 hemiplegic stroke patients and 30 healthy subjects. The

research's objective was two-fold. First, they wanted to explore the feasibility of using a foot contact pattern for neurologic recovery prediction. Second, they wanted to investigate the ambulation training effect in patients with hemiplegic stroke. The authors in Wang et al. [20] applied deep neural networks (DNNs) for stroke gait detection and abnormal gait pattern classification. The work in Jung et al. [21] created a functional electrical stimulation (FES) control system for enhancing gait rehabilitation in patients suffering from hemiplegia after a stroke. They used DNNs to control the muscles' electrical stimulation on the paretic side. The DNNs were trained using muscle activity data from healthy subjects during gait.

A significant percentage of studies used data taken from children suffering from hemiplegia, diplegia, or cerebral palsy. The authors in Di Nardo et al. [22] quantified the asymmetric behavior of children with mild hemiplegia during self-selected walking using features taken from surface-electromyography (EMG) and foot-floor contacts. They used statistical gait analysis in 16 hemiplegic and 100 healthy children. The study Aguilera and Subero [23] studied kinematic, kinetic, and EMG data from 278 children with spastic hemiplegia in order to find useful gait patterns. The authors in Abaid et al. [24] created a gait phase detection approach that is based on a hidden Markov model. It utilizes single-axis foot-mounted wearable gyroscopes to record the input data. The authors in Wang and Wang [25] conducted a gait analysis using 200 healthy and 20 children with spastic hemiplegic cerebral palsy. The authors in Zhang and Ma [26] compared the artificial neural network, discriminant analysis, naive Bayes, decision tree, k-nearest neighbors, SVM, and random forest (RF) machine learning algorithms in a gait classification system. The system contained gait data from 200 cerebral palsy children with spastic diplegia. This comparison revealed that the artificial neural network managed to achieve the highest precision accuracy. The work in Ferrari et al. [27] utilized a multi-layer perceptron (MLP) and a recursive neural network (RNN) to classify 174 diplegic children into the four walking pattern forms defined in the work of Ferrari et al. [28]. Kamruzzaman and Begg [29] applied SVM for the detection and classification of children having spastic diplegic gait in a dataset containing entries from 68 healthy and 88 non-healthy children. The system used stride length and cadence temporal-spatial gait parameters as an input feature vector to the SVM classifier. The authors in Orozco et al. [30] conducted an assessment analysis of gait abnormalities in 30 children with spastic cerebral palsy. The work in Strohrmann et al. [31] created a shoe-based wearable sensor system that utilizes active shape models to analyze the center of pressure trajectories for gait deviation assessment. The experimental part of their study involved the data collection from 15 children with cerebral palsy using supervised and semi-supervised settings.

Many studies were focused on recognizing multiple neurodegenerative diseases. The study Verlekar et al. [32] proposed a gait classification method for different diseases (diplegia, hemiplegia, neuropathy, and Parkinson's disease) combined with normal gait patterns. The system utilizes a pre-trained VGG-19 CNN for a generic image description. The authors in Pandit et al. [33] conducted a gait analysis using four wireless modules equipped with an accelerometer and gyroscope placed on the subject's body. The received signals were converted into image data and formed the input vector for transfer learning to the Inception v3 CNN. The CNN can distinguish between normal, hemiplegic, diplegic, and Parkinsonian gait. The study Nieto-Hidalgo and García-Chamizo [34] utilized a smartphone camera to record the sagittal view of the patient's gait. The recorded data were introduced as input to a computer vision algorithm for extracting different gait parameters. These parameters classified gait into five categories (normal, diplegic, hemiplegic, neuropathic, and Parkinsonian). The work in Goyal et al. [35] developed gait features from the human pose key points to detect Parkinson's disease, diplegia, hemiplegia, and Huntington's chorea. The authors in Nithyakani et al. [36] classified gait pathologies with the help of a CNN, which receives as input silhouette images taken from cameras. The study Pogorelc and Gams [37] developed a system for detecting abnormal gait patterns in older people which are related to health issues. The subject's gait is recorded using a system containing tags placed on the patient's body and sensors installed at their apartment. The model was tested using

the naive Bayes, and RF classification algorithms implementations in Weka [38,39]. It can recognize five gait types (normal, hemiplegic, Parkinson's disease, back pain, and leg pain), with the RF algorithm having better accuracy.

A number of systems are targeted for stroke patient rehabilitation. Zhang and Zhang [40] developed a hybrid network model for guiding the control strategy design of a robotic system for active rehabilitation training of stroke patients. The hybrid network combines a CNN and a spatial–temporal-embedded long–short-term memory (SQLSTM) model. The authors in Dose et al. [41] used a CNN on electroencephalogram (EEG) signals for the development of a brain-computer interface (BCI) system aimed at stroke patient rehabilitation. The study Buckley et al. [42] used a single wearable sensor on stroke patients' trunks to measure gait asymmetry, which is a targeted cardinal symptom during rehabilitation. They compared many acceleration-derived asymmetry variables utilizing an instrumented walkway (GaitRite). Their research concluded that step regularity was the most valid and reliable asymmetry outcome. The work in Kaku et al. [43] developed a method combining kinematic data from stroke patients wearing inertial measurement units with CNNs. The proposed system identifies functional primitives, which in turn form rehabilitation activities. Panwar et al. [44] created a framework based on deep learning that can classify three upper limb arm movements (extension, flexion, and forearm rotation). These movements can become a rehabilitation progress measure for stroke patients. The authors in Kashi et al. [45] developed a model that helps stroke patients during their rehabilitation. The model utilizes the RF algorithms and gives feedback to stroke patients regarding the undesirable movements they make during the execution of their rehabilitation exercises. The study Lin et al. [46] proposed a machine learning method for predicting daily living activities in stroke patients. All the above methods produce good results, but they cannot distinguish between the two hemiplegia types (right or left).

The proposed system was compared with the four classification algorithms described below. The study Broomhead and Lowe [47] created the radial basis function neural network (RBFNN) types which are inspired by the locally tuned response found in biological neurons. An RBFNN utilizes radial basis functions (RBFs) as neuron output functions, and its training process involves a one, two, or three-step manner [48]. Broyden, Fletcher, Goldfarb, and Shanno proposed a novel quasi-Newton method (BFGS) that utilizes a new updating formula for the Hessian's approximation [49]. Vapnik and Chervonenkis developed the support vector machine (SVM) method. Its main characteristic is the creation of an optimal or a set of optimal hyperplanes in a high or infinite-dimensional space. The created hyperplane(s) can be used in various tasks, including classification and regression [50,51]. Finally, the GenClass parallel tool can create classification programs in a C-like programming language using GE. It is designed to run in multi-core environments, which can significantly reduce its running time [52,53].

3. The FSC4RBF Algorithm

A genetic algorithm (GA) is a computational process inspired by Charles Darwin's theory of evolution. It solves a computation problem using an artificial evolution process that imitates biological procedures such as natural selection and survival of the fittest. A GA starts with the initial population, a collection of possible solutions to the problem. Each individual forming a possible solution is termed a chromosome and follows a specific encoding scheme. In early GA implementations, the bit-string encoding was adopted, where each individual was encoded as a bit-string stored in a vector or array. Alternative encoding schemes involve creating chromosomes having integer and real values. A GA uses the selection, crossover, and mutation procedures at each algorithm's iteration (generation). The selection process involves evaluating each individual and selecting the most promising ones for reproduction according to a fitness criterion. The crossover step reproduces the population by combining two existing chromosomes (parents) to produce a better solution (offspring). The initial crossover implementations with the bit-string encoding scheme utilized the one-point crossover. It involved creating offspring by using

the mutual exchange of information between two parent chromosomes. The mutation operator added diversity to the population by randomly selecting a position inside the chromosome (gene) and replacing it with another value from the same encoding scheme. The selection, crossover, and mutation procedures are repeated at each generation until the stopping criteria are satisfied [54,55].

The GE is a GA-based method utilizing an evolutionary process typically used to create computer programs. The term computer program denotes any instruction sequence that can perform a specific task ranging from a single expression to many statements. It is biologically inspired by the protein generation process and can generate a computer program in an arbitrary language with the help of a binary string with variable length. The structure of a GE contains a context-free grammar (CFG), a fitness function, and an evolutionary algorithm. The CFG describes the syntax of the created programs utilizing a rule set containing a start symbol, a collection of terminal symbols, non-terminal symbols, and production rules which are insensitive to the sentence's context. It uses a mapping process by utilizing the binary strings to select the production rules defined in a CFG using Backus–Naur form (BNF) metasyntax notation. The fitness function is used to evaluate the created programs' quality. Finally, the evolutionary algorithm (usually a GA) is selected to explore the search space of all the programs definable from the CFG for finding the program that best satisfies the fitness function. [56,57].

BNF is a meta-language used to specify the symbol sequences that define a syntactically valid program in a specific language and is characterized by the following five rules.

- The differentiation of non-terminal symbols from terminal letters is conducted using special brackets containing them.
- The definition alternatives are bundled together.
- Left is distinguished from right by adopting the ":=" symbol.
- The alternatives are distinguished with the "|" symbol.
- Full names describing the defined strings' meaning are utilized as non-terminal symbols [58,59].

In a CFG, the production rules have the form $V \rightarrow w$ where V denotes a non-terminal symbol, and w denotes a sequence comprised of terminal and non-terminal symbols. The mathematical definition of a CFG can be seen in Equation (1) where G defines the CNF, V_T, is a collection of terminal symbols, V_N, is a collection of non-terminal symbols, P are the production rules, and S denotes the start symbol [9].

$$G = (V_T, V_N, P, S) \tag{1}$$

In GE, a gene is an integer value defining a production rule from the P set. The chromosome creation follows an iterative procedure beginning from the start symbol, where the non-terminal symbols are replaced with the production rule's right part. The procedure follows the rule defined in Equation (2) where B denotes the next gene, and R_N are the production rules for the specific non-terminal symbol. The procedure continues until the end of the chromosome's length is reached. If an invalid expression is created, then the process is repeated from the chromosome's beginning, or the chromosome receives a low fitness value, and the mapping process terminates [9].

$$rule = B \bmod R_N \tag{2}$$

The FSC4RBF method is based on GE and uses an evolutionary method for the creation of new features with the purpose of lowering the generalization error in RBFNNs. Algorithm 1 describes the structure of the FSC4RBF approach [9].

Algorithm 1 The FSC4RBF Method

1: *create Population* $= [g_1, g_2, \ldots, g_{N_g}]$, $N_g \in \mathbb{N}$
Create the initial population.
2: **loop**
Start the evolution process.
3:      $f = [f_1, f_2, \ldots, f_{N_g}]$
Calculate each chromosome's fitness.
4:      **if** $(Gen_{No} = Gen_{Max})$ or $(best(Population) > solution_{acceptable})$ **then**
Check if the maximum number of generations is reached or if the chromosome
with the best fitness surpasses a predefined acceptable threshold.
5:          $solution_{best} \leftarrow best(Population)$
Select the best solution.
6:          **return** $solution_{best}$
End the evolution process and return the best solution.
7:      **end if**
8:      $Population_{select} \leftarrow select(Population(f))$
Create the new population from the fittest chromosomes.
9:      $Population_{crossover} \leftarrow crossover(Population_{select})$
Create the offspring.
10:     $Population \leftarrow mutation(Population_{crossover})$
Mutate the population.
11: **end loop**

The FSC4RBF method starts by creating the initial population (step 1), where each chromosome is encoded as a vector containing integer values (the total number of chromosomes in the initial population is denoted with the N_g symbol). The next step begins the evolution process by training a series of RBFNNs in a supervised manner where each network's fitness is calculated (step 3). The selected fitness criterion for classification problems is accuracy, while for regression problems it is the mean square error. In step 4, a condition checks if the maximum number of generations is reached or if the chromosome with the best fitness surpasses a predefined acceptable threshold. If the condition is satisfied, the best solution is selected (step 5) and returned (step 6). Then, the algorithm terminates. If the condition is not satisfied, the best chromosomes according to fitness are selected for reproduction using tournament selection (step 8). The reproduction process in step 9 utilizes the one-point crossover operator, which reproduces the population. The reproduction process from this operator is achieved by the mutual exchange of information between two parent chromosomes at a randomly chosen crossover point (Figure 1).
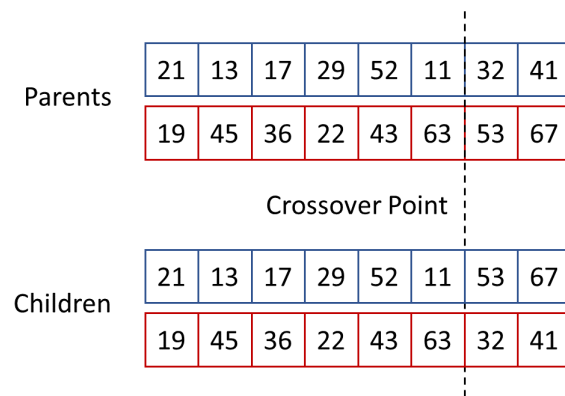
Figure 1. The one-point crossover operation. This figure presents the one-point reproduction operation where the information is mutually exchanged in a randomly chosen crossover point between two parent chromosomes.

The offspring creation process follows Equation (3) where n is the number of offspring, R_s is the number of parent chromosomes that will remain unchanged in the next generation, and N_g is the number of chromosomes in the population.

$$n = (1 - R_s)N_g \tag{3}$$

Step 10 mutates the population. It utilizes the fixed mutation rate R_m and every gene is assigned a random number $r \in [0, 1]$. If the assigned random number is lower or equal to a predefined threshold P_M, the gene's value is randomly changed. Finally, step 11 continues the evolutionary procedure [9].

4. System Architecture

The proposed system uses the RehaGait mobile gait analysis system [6]. This system has seven sensors that are placed in various patient body parts. Each sensor is equipped with an accelerometer, a magnetometer, and a gyroscope. Only the signals taken from the accelerometer are utilized, and they are windowed into 2-second windows before undergoing a pre-processing and feature extraction stage. The RehaGait mobile gait analysis system is equipped with a wireless module that allows signal transmission to a personal computer (PC) without a data cable. The pre-processing procedure involves the adoption of a low-pass filter to smoothen the captured signals. The low-pass filter contains a 6 Hz pass-band frequency, and each captured signal was sampled at 1000 Hz.

The feature extraction stage utilizes four time-domain and two frequency-domain features. The time-domain features are the mean (μ), the standard deviation (S), the kurtosis (k), and the peak-magnitude-to-RMS ratio (RMS). The mean can be seen in Equation (4), where N is the number of scalar observations, and A defines a random variable vector.

$$\mu = \frac{1}{N} \sum_{i=1}^{N} A_i \tag{4}$$

The standard deviation is depicted in Equation (5) where N defines the number of scalar observations, A determines a random variable vector, and μ is its mean.

$$S = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (A_i - \mu)^2} \tag{5}$$

A distribution's kurtosis is defined in Equation (6) where $E(t)$ is the expected quantity value t, μ defines the mean of x and σ denotes the standard deviation of x.

$$k = \frac{E(x - \mu)^4}{\sigma^4} \tag{6}$$

The peak-magnitude-to-RMS ratio is defined in Equation (7).

$$RMS = \frac{\|A\|_\infty}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} |A_i|^2}} \tag{7}$$

The frequency-domain features are the acceleration energy and the acceleration signal energy. The acceleration energy is shown in Equation (8) where A_i is the acceleration signal's ith spectral line and N defines the total number of lines.

$$Eng = \frac{\sum_{i=1}^{N} A_i^2}{N} \tag{8}$$

The acceleration signal entropy completes the frequency-domain feature selection. It is defined in Equation (9) where p_i is the probability of A_i to occur in the amplitude spectrum.

$$Ent = -\sum_{i=1}^{N} p_i \log_2 p_i \tag{9}$$

The combination of time-domain and frequency-domain features forms the input vector to the FSC4RBF method, which creates optimized features that increase the classification accuracy of the RBF network. The RBF network classified the data taken from a custom-made dataset into three classes (healthy, left, or right hemiplegia). The system architecture can be seen in Figure 2.
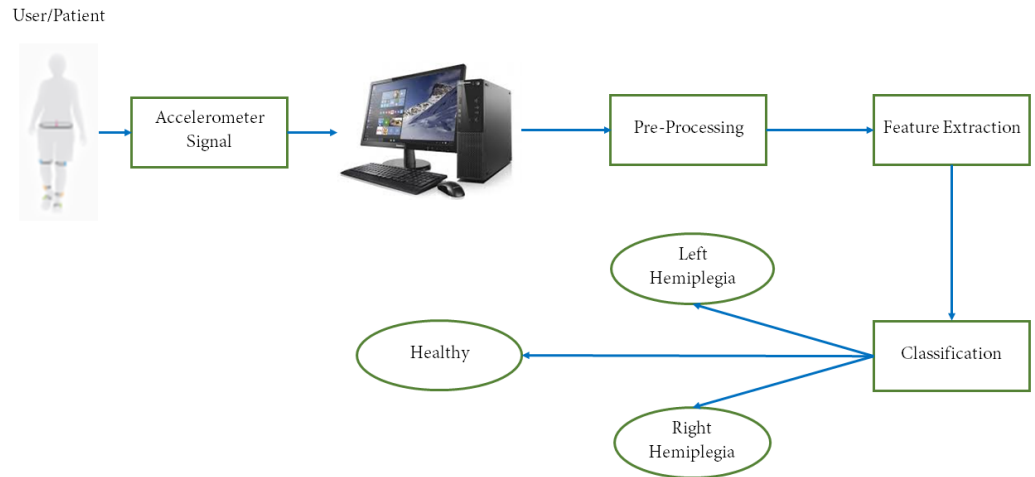
Figure 2. The system architecture. The system utilizes seven sensors placed in various body parts of the patient. The data from the accelerometer signals are received by a PC using a wireless connection and undergo a pre-processing and feature extraction procedure. Then, they introduced it as input to the FSC4RBF method, which creates optimized features that increase the classifier's accuracy (RBFNN). Finally, the classifier categorizes the input data into three classes (healthy, left, or right hemiplegia).

Visualization of three input patterns after the pre-processing and manual feature extraction stages are applied to the accelerometer's signals dataset is depicted in Figure 3. These input patterns correspond to three signal types (normal, left, and right hemiplegia).
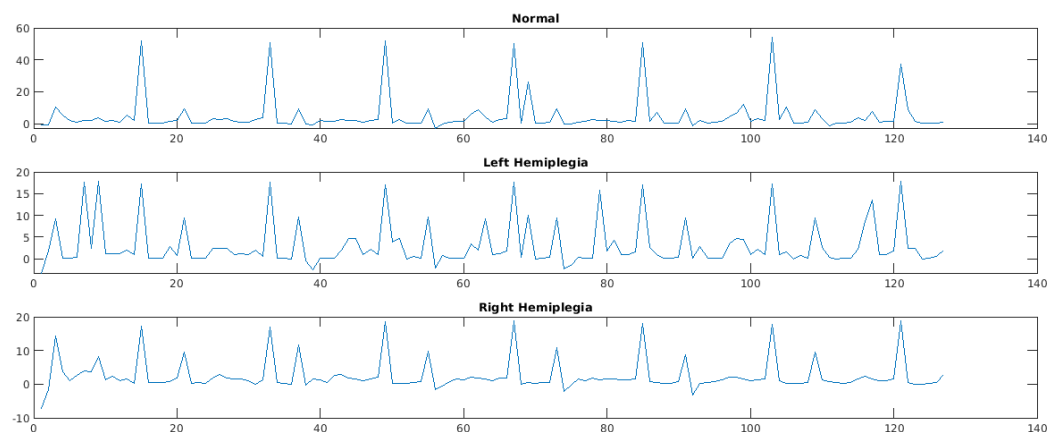
Figure 3. A visualization of three input patterns. The current figure depicts three input patterns after the pre-processing and feature extraction stages are applied to the accelerometer's signals dataset corresponding to three signal types (normal, left, and right hemiplegia).

5. Experimental Results

The FSC4RBF method was tested with four classification approaches (RBFNN, BFGS, SVM, and GenClass) in a custom-created dataset containing entries taken from ten healthy and twenty hemiplegic patients (eight had left hemiplegia and twelve had right hemiplegia). The data from the seven accelerometers were sent to a PC which divided them into 2-s windows. The signals were pre-processed with the help of a low-pass filter before undergoing a feature extraction stage using four time-domain and two frequency-domain features. The dataset contains the values after the pre-processing and the initial manual feature extraction stage. All input feature entries are real-valued. The low-pass filter contains a 6 Hz pass-band frequency, and each captured signal was sampled at 1000 Hz. The above features formed the input vector to the FSC4RBF method and constituted the initial population of the modified GE algorithm. The modified GE algorithm automatically evolved these initial features into more advanced ones, which were used to train an RBFNN.

5.1. Parameter Details

The experiments were run 30 times in all neural network-based methods to avoid any bias toward the random initialization of hidden nodes and thresholds, while the number of hidden layer neurons (h) was fixed at 10 nodes. The dataset was divided into training/test sets using 10-fold cross-validation. The two evolutionary approaches (GenClass and FSC4RBF) were run for 200 generations ($Gens_{No}$) with a population size (N_g) of 500 individuals. Both methods utilized the one-point crossover operator and adopted a fixed 5% mutation rate (R_m). The parameters used in all the experiments are depicted in Table 1. Finally, the same pre-processing and initial manual feature extraction stage was applied in all experiment runs to avoid any bias between the compared methods.

Table 1. ParameterSettings.

Parameter Name	Symbol	Values/Types
Experiments Repeats for Neural Network Methods	*expNo*	30
Number of Hidden Layer Nodes	*h*	10
Number of Folds	*k*	10
Generations No	$Gens_{No}$	200
Population Size	N_g	500
Crossover	*crossover*	one-point
Mutation Rate	R_m	5%

5.2. Comparison Results

The selected statistic for comparing the FSC4RBF method with four classification methods was the classification accuracy (*acc*) defined in Equation (10). The term k in the accuracy formula defines the number of folds, *err* is the number of erroneously classified test patterns, and p denotes the number of test patterns introduced to the system.

$$acc = \frac{1}{k}\sum_{i=1}^{k}\left(1 - \frac{err}{p}\right). \tag{10}$$

The outcome from the comparison of the five classification methods is summarized in Table 2. The results show that the proposed FSC4RBF method achieved the highest classification accuracy compared to the other four approaches.

Table 2. Results Comparison.

RBFNN	BFGS	SVM	GenClass	FSC4RBF
74.53%	83.71%	80.69%	87.98%	90.07%

The results are also visualized in the bar chart in Figure 4.
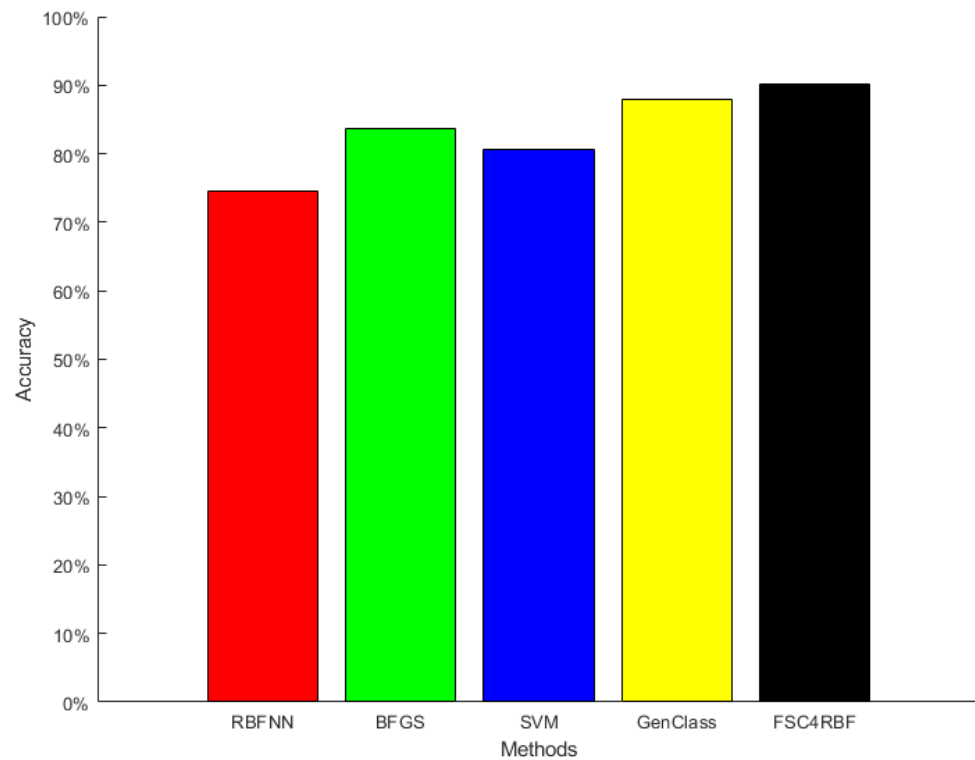
Figure 4. Comparison results plot. This figure visualizes the classification accuracy between the proposed FSC4RBF algorithm and four existing methods. It is shown that the proposed method managed to achieve the highest accuracy (black bar) in the custom-created dataset.

6. Discussion

The proposed FSC4RBF system can be a helpful supporting tool for doctors or other specialists to diagnose the hemiplegia type between healthy and non-healthy individuals playing a significant role in their rehabilitation plan. The experimental results presented in Section 5.2 validated that the optimized features created from the FSC4RBF algorithm significantly increased the classification accuracy of RBFNN. The RBFNN using the initial features had only a 74.53% classification accuracy, the lowest compared to FSC4RBF and the other three classifiers. The results in Table 2 revealed that the BFGS, SVM, GenClass, and FSC4RBF algorithms had an average of 9.18%, 6.16%, 13.45%, and 15.54% performance increase in their test set results compared to the RBFNN without feature construction.

All methods were run one additional time to create their confusion matrices. In each confusion matrix, the green boxes denote the samples (bold fonts) and their percentages which are classified correctly. In contrast, the misclassified samples (bold fonts) and their percentages are visualized with red boxes. The vertical white boxes show the percentages of all the samples predicted to belong to each class classified correctly (green color) and incorrectly (red color). The horizontal white boxes show the percentages of all the examples belonging to each class and are classified correctly (green color) and incorrectly (red color). Finally, the overall accuracy is visualized with a gray box. The confusion matrix of the FC4RBF method is depicted in Figure 5. It is shown that FC4RBF had the highest classification accuracy (92%) in the right hemiplegia class, while the lowest (76%) was in the normal class.

Figure 5. The FC44RBF confusion matrix. The current figure depicts the confusion matrix of the proposed FC4RBF method.

Similarly, in the four confusion matrices of the compared methods in Figure 6, the highest accuracies were achieved in the right hemiplegia class, while the lowest ones were observed in the normal class. One exception was the SVM method, where the lowest accuracy was observed in the left hemiplegia class.

Creating optimized features using a hybrid approach combining an RBFNN with a GE algorithm resulted in the RBFNN's classification accuracy increase. The main issue of using this approach was the significant computational overhead introduced by the GE method, which required many networks to be trained at each generation. The solution to this problem can be adopting a parallel processing scheme using a multi-core central processing unit at each algorithm's iteration.

The experimental results from the proposed FSC4RBF system were compared with existing approaches for hemiplegic gait detection. The hemiplegic gait detection method by Lee et al. [8] utilized the random forest (RF) algorithm for the classification task between hemiplegic and normal gait. The RF method managed to obtain 100% classification accuracy. The authors in Patil et al. [10] tested one shallow and three deep architectures for detecting hemiplegic gait. The deep architectures were based on visual geometry group (VGG)-16 CNN. The best accuracy (99.68% on the validation set) was achieved from the third DNN. Yardimci [11] got a 100% classification accuracy using a fuzzy logic-based method. Luo et al. [15] used the same algorithm with different combinations of extracted features achieving a 90.65% average accuracy between all gait feature combinations and a 95.45% best accuracy. The study Wang et al. [20] used DNN architectures to detect stroke gaits and classify abnormal gait patterns. The created models achieved a 99.35% classification accuracy in detecting stroke gaits. The above methods managed to obtain impressive results, with some even reaching a 100% classification accuracy compared to the 90.07% of the proposed FSC4RBF method. The main reason for this difference is that they are not distinguishing between the left and right hemiplegia types. Still, they are trying to solve a simpler binary classification problem containing only hemiplegic and normal individuals.
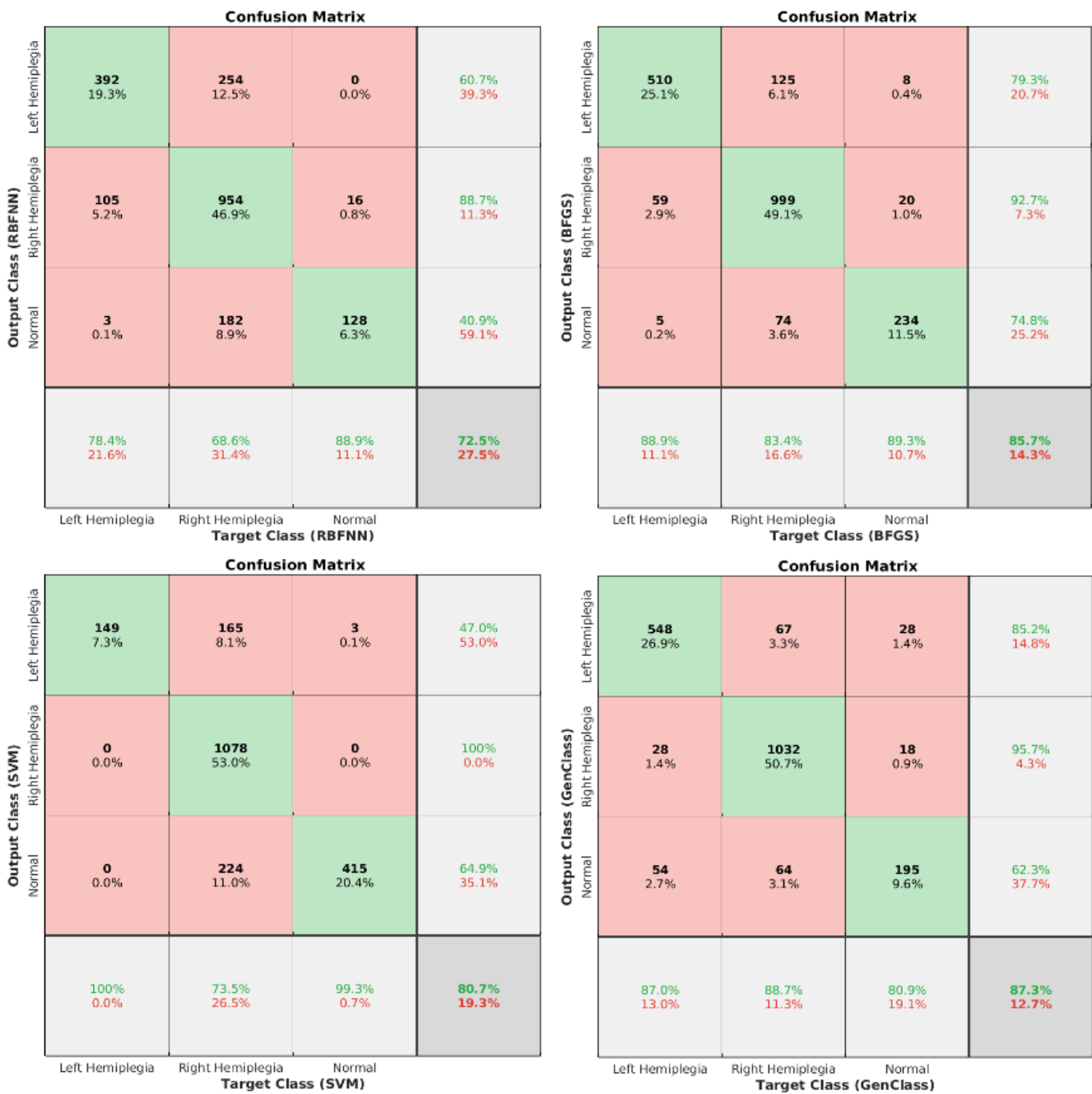
Figure 6. The compared methods confusion matrices. The current figure depicts the confusion matrices for all compared methods (RBFNN, BFGS, SVM, GenClass).

7. Conclusions

This paper introduced a system for automatically classifying hemiplegia type (right or left) between healthy and non-healthy subjects. The system utilized a feature construction method for RBF neural networks (FSC4RBF). The FSC4RBF method created optimized complex features from six simple custom-created ones (four time-domain and two frequency-domain features). The proposed system achieved the highest classification accuracy and revealed that using an evolutionary approach to create novel, optimized features from existing ones can significantly increase the classification accuracy of the network.

The FC4RBF method confusion matrix in the previous section revealed an increased number of misclassified healthy classes as right hemiplegic ones. One possible explanation for this finding would be the class imbalance in the dataset (ten healthy, eight patients with left hemiplegia, and twelve with right hemiplegia).

Future research involves expanding the dataset to include more hemiplegic patients, further increasing the classification accuracy. Another future work includes other disease types such as diplegia and Parkinson's disease.

Author Contributions: Conceptualization, I.T., E.G. and N.G.; methodology, V.C., A.A. and I.T.; software, I.T. and A.A.; validation, A.T.T. and M.G.T.; investigation, C.G. and A.T.T.; data curation, D.V., D.D. and A.P.; writing—original draft preparation, V.C., A.A., D.D., D.V. and I.T.; writing—review and editing, I.T., A.T.T., C.G., M.G.T., E.G., A.P. and N.G.; visualization, V.C.; supervision, A.P. and N.G.; funding acquisition, A.P. and N.G. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by Ethics Committee of University of Ioannina.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patient(s) to publish this paper.

Data Availability Statement: New data were created and analyzed in this study. Data sharing not applicable.

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

References

1. Davies, P.M. *Right in the Middle: Selective Trunk Activity in the Treatment of Adult Hemiplegia*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1990.
2. Bobath, B. *Adult Hemiplegia: Evaluation and Treatment*, 3 ed.; Butterworth-Heinemann: Oxford, UK, 1990.
3. Ruskin, A. Understanding stroke and its rehabilitation. *Stroke* **1983**, *14*, 438–442. [CrossRef] [PubMed]
4. Davies, P.M. *Steps to Follow: The Comprehensive Treatment of Patients with Hemiplegia*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2000.
5. Donath, L.; Faude, O.; Lichtenstein, E.; Pagenstert, G.; Nüesch, C.; Mündermann, A. Mobile inertial sensor based gait analysis: Validity and reliability of spatiotemporal gait characteristics in healthy seniors. *Gait Posture* **2016**, *49*, 371–374. [CrossRef] [PubMed]
6. HASOMED. RehaGait—Mobile Gait Analysis. 2022. Available online: https://hasomed.de/en/products/rehagait/ (accessed on 6 May 2022).
7. Schwesig, R.; Fischer, D.; Lauenroth, A.; Becker, S.; Leuchte, S. Can falls be predicted with gait analytical and posturographic measurement systems? A prospective follow-up study in a nursing home population. *Clin. Rehabil.* **2013**, *27*, 183–190. [CrossRef] [PubMed]
8. Lee, J.; Park, S.; Shin, H. Detection of hemiplegic walking using a wearable inertia sensing device. *Sensors* **2018**, *18*, 1736. [CrossRef]
9. Gavrilis, D.; Tsoulos, I.G.; Dermatas, E. Selecting and constructing features using grammatical evolution. *Pattern Recognit. Lett.* **2008**, *29*, 1358–1365. [CrossRef]
10. Patil, S.; Shah, A.; Dalvi, S.; Sisodia, J. Early detection of hemiplegia by analyzing the gait characteristics and walking patterns using convolutional neural networks. In Proceedings of the International Conference on Soft Computing and Signal Processing, Dalian, China, 20–23 September 2019; Springer: Singapore, 2019; pp. 39–51.
11. Yardimci, A. Fuzzy logic based gait classification for hemiplegic patients. In Proceedings of the International Symposium on Intelligent Data Analysis, Ljubljana, Slovenia, 6–8 September 2007; Springer: Berlin/Heidelberg, Germany, 2007; pp. 344–354.
12. Manca, M.; Ferraresi, G.; Cosma, M.; Cavazzuti, L.; Morelli, M.; Benedetti, M. Gait patterns in hemiplegic patients with equinus foot deformity. *BioMed Res. Int.* **2014**, *2014*, 939316. [CrossRef]
13. Mulroy, S.; Gronley, J.; Weiss, W.; Newsam, C.; Perry, J. Use of cluster analysis for gait pattern classification of patients in the early and late recovery phases following stroke. *Gait Posture* **2003**, *18*, 114–125. [CrossRef]
14. Straudi, S.; Manca, M.; Aiello, E.; Ferraresi, G.; Cavazza, S.; Basaglia, N. Sagittal plane kinematic analysis of the six-minute walk test: A classification of hemiplegic gait. *Eur. J. Phys. Rehabil. Med.* **2009**, *45*, 341–347.
15. Luo, G.; Zhu, Y.; Wang, R.; Tong, Y.; Lu, W.; Wang, H. Random forest–based classsification and analysis of hemiplegia gait using low-cost depth cameras. *Med. Biol. Eng. Comput.* **2020**, *58*, 373–382. [CrossRef]

16. Zhu, Y.; Xu, W.; Wang, R.; Tong, Y.; Lu, W.; Wang, H. Automatic recognition and analysis of hemiplegia gait. *Sheng wu yi xue Gong Cheng xue za zhi= J. Biomed. Eng.= Shengwu Yixue Gongchengxue Zazhi* **2019**, *36*, 306–314.

17. Pauk, J.; Minta-Bielecka, K. Gait patterns classification based on cluster and bicluster analysis. *Biocybern. Biomed. Eng.* **2016**, *36*, 391–396. [CrossRef]

18. Wang, H.; Wang, X.; Zhang, X.; Qiu, C.; Li, Q.; Zhao, T. Classification of Stroke Disease Assessment based on Body Surface Electrical Signals at Acupuncture Points. *J. Phys.* **2021**, *1924*, 012022. [CrossRef]

19. Wong, A.M.; Pei, Y.C.; Hong, W.H.; Chung, C.Y.; Lau, Y.C.; Chen, C.P. Foot contact pattern analysis in hemiplegic stroke patients: An implication for neurologic status determination. *Arch. Phys. Med. Rehabil.* **2004**, *85*, 1625–1630. [CrossRef] [PubMed]

20. Wang, F.C.; Chen, S.F.; Lin, C.H.; Shih, C.J.; Lin, A.C.; Yuan, W.; Li, Y.C.; Kuo, T.Y. Detection and Classification of Stroke Gaits by Deep Neural Networks Employing Inertial Measurement Units. *Sensors* **2021**, *21*, 1864. [CrossRef] [PubMed]

21. Jung, S.; Bong, J.H.; Kim, S.J.; Park, S. DNN-Based FES Control for Gait Rehabilitation of Hemiplegic Patients. *Appl. Sci.* **2021**, *11*, 3163. [CrossRef]

22. Di Nardo, F.; Strazza, A.; Mengarelli, A.; Cardarelli, S.; Tigrini, A.; Verdini, F.; Nascimbeni, A.; Agostini, V.; Knaflitz, M.; Fioretti, S. EMG-based characterization of walking asymmetry in children with mild hemiplegic cerebral palsy. *Biosensors* **2019**, *9*, 82. [CrossRef] [PubMed]

23. Aguilera, A.; Subero, A. Automatic gait classification patterns in spastic hemiplegia. *Adv. Data Anal. Classif.* **2020**, *14*, 897–925. [CrossRef]

24. Abaid, N.; Cappa, P.; Palermo, E.; Petrarca, M.; Porfiri, M. Gait detection in children with and without hemiplegia using single-axis wearable gyroscopes. *PLoS ONE* **2013**, *8*, e73152. [CrossRef]

25. Wang, X.; Wang, Y. Gait analysis of children with spastic hemiplegic cerebral palsy. *Neural Regen. Res.* **2012**, *7*, 1578.

26. Zhang, Y.; Ma, Y. Application of supervised machine learning algorithms in the classification of sagittal gait patterns of cerebral palsy children with spastic diplegia. *Comput. Biol. Med.* **2019**, *106*, 33–39. [CrossRef]

27. Ferrari, A.; Bergamini, L.; Guerzoni, G.; Calderara, S.; Bicocchi, N.; Vitetta, G.; Borghi, C.; Neviani, R.; Ferrari, A. Gait-based diplegia classification using lsmt networks. *J. Healthc. Eng.* **2019**, *2019*, 3796898. [CrossRef]

28. Ferrari, A.; Alboresi, S.; Muzzini, S.; Pascale, R.; Perazza, S.; Cioni, G. The term diplegia should be enhanced. Part I: A new rehabilitation oriented classification of cerebral palsy. *Eur. J. Phys. Rehabil. Med.* **2008**, *44*, 195.

29. Kamruzzaman, J.; Begg, R.K. Support vector machines and other pattern recognition approaches to the diagnosis of cerebral palsy gait. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 2479–2490. [CrossRef]

30. Orozco, M.d.P.D.; Abousamra, O.; Church, C.; Lennon, N.; Henley, J.; Rogers, K.J.; Sees, J.P.; Connor, J.; Miller, F. Reliability and validity of Edinburgh visual gait score as an evaluation tool for children with cerebral palsy. *Gait Posture* **2016**, *49*, 14–18. [CrossRef] [PubMed]

31. Strohrmann, C.; Patel, S.; Mancinelli, C.; Deming, L.C.; Chu, J.J.; Greenwald, R.; Tröster, G.; Bonato, P. Automated assessment of gait deviations in children with cerebral palsy using a sensorized shoe and active shape models. In Proceedings of the 2013 IEEE International Conference on Body Sensor Networks, Cambridge, MA, USA, 6–9 May 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 1–6.

32. Verlekar, T.T.; Correia, P.L.; Soares, L.D. Using transfer learning for classification of gait pathologies. In Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 3–6 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 2376–2381.

33. Pandit, T.; Nahane, H.; Lade, D.; Rao, V. Abnormal gait detection by classifying inertial sensor data using transfer learning. In Proceedings of the 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA), Boca Raton, FL, USA, 16–19 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1444–1447.

34. Nieto-Hidalgo, M.; García-Chamizo, J.M. Classification of Pathologies Using a Vision Based Feature Extraction. In Proceedings of the Ubiquitous Computing and Ambient Intelligence, Philadelphia, PA, USA, 7–10 November 2017; Ochoa, S.F., Singh, P., Bravo, J., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 265–274.

35. Goyal, D.; Jerripothula, K.R.; Mittal, A. Detection of gait abnormalities caused by neurological disorders. In Proceedings of the 2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP), Tampere, Finland, 21–24 September 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.

36. Nithyakani, P. Classification Of Gait Pathology Using Enhanced Convolutional Neural Network. In Proceedings of the 2022 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 25–27 January 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–5.

37. Pogorelc, B.; Gams, M. Identification of gait patterns related to health problems of elderly. In Proceedings of the International Conference on Ubiquitous Intelligence and Computing, Xi'an China, 26–29 October 2010; Springer: Berlin/Heidelberg, Germany, 2010; pp. 179–191.

38. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The WEKA data mining software: An update. *ACM SIGKDD Explor. Newsl.* **2009**, *11*, 10–18. [CrossRef]

39. Frank, E.; Hall, M.A.; Witten, I.H. *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed.; Morgan Kaufmann: Burlington, MA, USA, 2016.

40. Zhang, P.; Zhang, J. Deep learning analysis based on multi-sensor fusion data for hemiplegia rehabilitation training system for stoke patients. *Robotica* **2022**, *40*, 780–797. [CrossRef]

41. Dose, H.; Møller, J.S.; Iversen, H.K.; Puthusserypady, S. An end-to-end deep learning approach to MI-EEG signal classification for BCIs. *Expert Syst. Appl.* **2018**, *114*, 532–542. [CrossRef]

42. Buckley, C.; Micó-Amigo, M.E.; Dunne-Willows, M.; Godfrey, A.; Hickey, A.; Lord, S.; Rochester, L.; Del Din, S.; Moore, S.A. Gait asymmetry post-stroke: Determining valid and reliable methods using a single accelerometer located on the trunk. *Sensors* **2019**, *20*, 37. [CrossRef]

43. Kaku, A.; Parnandi, A.; Venkatesan, A.; Pandit, N.; Schambra, H.; Fernandez-Granda, C. Towards data-driven stroke rehabilitation via wearable sensors and deep learning. *Proc. Mach. Learn. Healthc. Conf. PMLR* **2020**, *126*, 143–171.

44. Panwar, M.; Biswas, D.; Bajaj, H.; Jöbges, M.; Turk, R.; Maharatna, K.; Acharyya, A. Rehab-net: Deep learning framework for arm movement classification using wearable sensors for stroke rehabilitation. *IEEE Trans. Biomed. Eng.* **2019**, *66*, 3026–3037. [CrossRef] [PubMed]

45. Kashi, S.; Polak, R.F.; Lerner, B.; Rokach, L.; Levy-Tzedek, S. A machine-learning model for automatic detection of movement compensations in stroke patients. *IEEE Trans. Emerg. Top. Comput.* **2020**, *9*, 1234–1247. [CrossRef]

46. Lin, W.Y.; Chen, C.H.; Tseng, Y.J.; Tsai, Y.T.; Chang, C.Y.; Wang, H.Y.; Chen, C.K. Predicting post-stroke activities of daily living through a machine learning-based approach on initiating rehabilitation. *Int. J. Med. Inform.* **2018**, *111*, 159–164. [CrossRef] [PubMed]

47. Broomhead, D.S.; Lowe, D. *Radial Basis Functions, Multi-Variable Functional Interpolation and Adaptive Networks*; Technical Report; Royal Signals and Radar Establishment: Malvern, UK, 1988.

48. Schwenker, F.; Kestler, H.A.; Palm, G. Three learning phases for radial-basis-function networks. *Neural Netw.* **2001**, *14*, 439–458. [CrossRef]

49. Hery, M.A.; Ibrahim, M.; June, L. BFGS method: A new search direction. *Sains Malays.* **2014**, *43*, 1591–1597.

50. Zahir, N.; Mahdi, H. Snow depth estimation using time series passive microwave imagery via genetically support vector regression (case study urmia lake basin). *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *40*, 555. [CrossRef]

51. Suthaharan, S. Support Vector Machine. In *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*; Springer: Boston, MA, USA, 2016; pp. 207–235. [CrossRef]

52. Tsoulos, I.G. Creating classification rules using grammatical evolution. *Int. J. Comput. Intell. Stud.* **2020**, *9*, 161–171.

53. Anastasopoulos, N.; Tsoulos, I.G.; Tzallas, A. GenClass: A parallel tool for data classification based on Grammatical Evolution. *SoftwareX* **2021**, *16*, 100830. [CrossRef]

54. Soricone, R.; Neville, M. Comparative analysis of genetic algorithm implementations. *ACM SIGAda Ada Lett.* **2004**, *24*, 35–38. [CrossRef]

55. Kramer, O. Genetic algorithms. In *Genetic Algorithm Essentials*; Springer: Cham, Switzerland, 2017; pp. 11–19.

56. O'Neill, M.; Ryan, C. Grammatical evolution. *IEEE Trans. Evol. Comput.* **2001**, *5*, 349–358. [CrossRef]

57. Noorian, F.; de Silva, A.M.; Leong, P.H. *Grammatical Evolution: A Tutorial Using gramEvol*; Massachusetts Institute of Technology: Cambridge, MA, USA, 2016.

58. Knuth, D.E. Backus normal form vs. backus naur form. *Commun. ACM* **1964**, *7*, 735–736. [CrossRef]

59. McCracken, D.D.; Reilly, E.D. Backus-naur form (bnf). In *Encyclopedia of Computer Science*; John Wiley and Sons Ltd.: Chichester, UK, 2003; pp. 129–131.