Artificial intelligence to improve efficiency of administration of gross motor function assessment in children with cerebral palsy

IBRAHIM DURAN | CHRISTINA STARK | AHMET SAGLAM | ALEXANDRA SEMMELWEIS | HEIDRUN LIOBA WUNRAM | KAROLINE SPIESS | ECKHARD SCHOENAU

1 Center of Prevention and Rehabilitation, Medical Faculty and University Hospital, University of Cologne, Cologne, Germany; 2 Department of Neurology, Medical Faculty and University Hospital, University of Cologne, Cologne, Germany; 3 Merzifon Vocational School Computer Technologies Department, University of Amasya, Amasya, Turkey; 4 Department of Psychiatry, Psychosomatics and Psychotherapy for Children and Adolescents, Medical Faculty and University Hospital, University of Cologne, Cologne; 5 Department of Pediatrics, Medical Faculty and University Hospital, University of Cologne, Cologne, Germany.

Correspondence to Ibrahim Duran, Center of Prevention and Rehabilitation, University of Cologne, UniReha Lindenburger Allee 44, 50931 Cologne, Germany. E-mail: ibrahim.duran@unireha-koeln.de

PUBLICATION DATA
Accepted for publication 15th July 2021.
Published online 13th August 2021.

ABBREVIATIONS
FNN Feed forward neural net
GMFM Gross Motor Function Measure
ICC Intraclass correlation coefficient
MAE Mean absolute error
MSE Mean squared error
SVM Support vector machine

AIM To create a reduced version of the 66-item Gross Motor Function Measure (rGMFM-66) using innovative artificial intelligence methods to improve efficiency of administration of the GMFM-66.

METHOD This study was undertaken using information from an existing data set of children with cerebral palsy participating in a rehabilitation programme. Different self-learning approaches (random forest, support vector machine [SVM], and artificial neural network) were evaluated to estimate the GMFM-66 score with the fewest possible test items. Test agreements were evaluated (among other statistics) by intraclass correlation coefficients (ICCs).

RESULTS Overall, 1217 GMFM-66 assessments (509 females, mean age 8y 10mo [SD 3y 9mo]) at a single time and 187 GMFM-66 assessments and reassessments (80 females, mean age 8y 5mo [SD 3y 10mo]) after 1 year were evaluated. The model with SVM predicted the GMFM-66 scores most accurately. The ICCs of the rGMFM-66 and the full GMFM-66 were 0.997 (95% confidence interval [CI] 0.996–0.997) at a single time and 0.993 (95% CI 0.993–0.995) for the evaluation of the change over time.

INTERPRETATION The study shows that the efficiency of the full GMFM-66 assessment can be increased by using machine learning (self-learning algorithms). The presented rGMFM-66 score showed an excellent agreement with the full GMFM-66 score when applied to a single assessment and when evaluating the change over time.

Over the past decade, the efficacy of motor interventions for children with mobility disorders such as cerebral palsy (CP) has frequently been investigated. Nowadays, the identification of responder and non-responder is becoming increasingly important for quality management in rehabilitation.1 In this sense, a responder is someone who responds to a medical treatment with a clinically relevant (usually positive) effect. In addition, the identification of useful therapies and the avoidance of harmful therapies are important for patients and their families.

The Gross Motor Function Measure (GMFM) is a commonly used assessment tool designed and evaluated to measure changes in gross motor function over time in children and adolescents with CP.2 The original version of the GMFM contains 88 test items (GMFM-88). Studies have shown excellent interrater and test–retest reliability, and internal consistency in children with CP, but the GMFM-88 score had only an ordinal scale level.3 To improve its effectiveness, a Rasch analysis was performed and a 66 test-item version (GMFM-66) was developed with comparable reliability and validity.4 The GMFM-66 provides a total score based on an interval scale, which has clear advantages for mathematical processing, especially when used as an outcome parameter in clinical settings as well as in studies.

The entire test of the GMFM-66 takes between 45 minutes and 60 minutes for someone familiar with the measure, which is often an obstacle for administration in clinical practice.5 To further improve the efficiency of administration of the GMFM-66, the item-set method (GMFM-66-IS) and the GMFM-66 basal and ceiling (GMFM-66 B&C) were constructed.6 The abbreviated versions of the GMFM-66 have been developed to facilitate the best choice of test items. The GMFM-66-IS takes approximately 20 to 30 minutes to perform.

Artificial intelligence-powered medical technologies are increasingly used in clinical practice.6,7 They are based, for example, on self-learning algorithms which could be used...
for classification and regression tasks in medicine. Self-learning algorithms are computer programs that analyse the relation between the input (here: GMFM-66 items) and the output (here: the GMFM-66 score) on the basis of example input–output pairs to predict new, unknown input data. This kind of algorithm is also called supervised machine learning and is often used to create artificial intelligence. Self-learning algorithms can also be used to reduce the required features for creating a statistical model (feature selection).

The aim of this study was to create a reduced version of the GMFM-66 (rGMFM-66) using innovative artificial intelligence methods to improve efficiency of administration of the GMFM-66.

METHOD
The present study was a single-centre retrospective analysis of prospectively collected data of children with CP, who participated in the rehabilitation programme ‘Auf die Beine’ (‘On your feet’) at the Centre of Prevention and Rehabilitation (University of Cologne, Germany) from January 2006 to January 2020. The rehabilitation programme is financed as part of the regular rehabilitation programme of the health care system.

The rehabilitation programme combines intensive, goal-directed training during two short inpatient stays and continuous, intensive whole-body vibration training at home for 6 months. Assessments were applied at baseline (M0), after 6 months (M6) of additional training, and after 12 months (M12) without additional training (6mo of follow-up).

The inclusion criteria for this study were a main diagnosis of CP and age below 18 years. After receiving written informed consent from the legal representative of the child, clinical data were stored in a prospective single-centre patient registry. The Ethics Committee of the University of Cologne approved this registry (16-269). A detailed description of the registry can be found at www.germanc.tr.de (DRKS0001131).

Assessment of motor function by the GMFM-66
The GMFM-66 is an observational, clinical measure to evaluate gross motor function in children with CP. It is used routinely within the rehabilitation programme ‘Auf die Beine’ at the start, after 6 months, and after 12 months. It consists of 66 motor tasks (items), is validated, and is commonly used to quantify motor skills in children with CP (maximum score 100 points). The results of the individual items were analysed using the Gross Motor Ability Estimator (version 2) scoring software for the GMFM (CanChild, McMaster University, Ontario, Canada).

Classification of motor function by the GMFCS
The Gross Motor Function Classification System (GMFCS) was used to classify the children according to their motor function level. The GMFCS consists of a 5-point ordinal scale, designated as I to V. Children classified in GMFCS levels I and II can walk without support or with limitations respectively; children in GMFCS levels III and IV can only walk using handheld mobility devices or powered mobility. Children are transported in a manual wheelchair in all settings. Children are limited in their ability to maintain antigravity head and trunk postures and control leg and arm movements.

Study population
From January 2006 to January 2020, 5747 GMFM-66 measurements were recorded (Fig. S1, online supporting information). These included longitudinal measurements of the same patient 6 months apart. Two hundred and seventeen measurements were excluded, because they did not meet the age inclusion criteria. In clinical routine, it is not always possible to collect 100% complete GMFM-66 measurements, because sometimes the child is tired or other external or child-related factors make full testing impossible. Additionally, for internal quality management, we decided to omit item 46 (crawls up four steps on hands and knees/feet) if the child/adolescent felt too exposed in this task. For those reasons, many (n=4312) of the GMFM-66 assessments had to be excluded, because we only wanted to include 100% complete measurements for validation reasons. Nevertheless, 1217 measurements of the complete GMFM-66 could be included in the analysis. These GMFM-66 assessments were used for training, tuning, and validating the statistical models (see below and Fig. S2, online supporting information).

Additionally, there were 187 children with complete GMFM-66 assessments at M0 and M12. The change over time was defined as the GMFM-66 score at M12 minus the GMFM-66 score at M0. The data of these 187 children were used to compare the change over time using the rGMFM-66 and the GMFM-66.

Table 1 presents the demographics of the study populations. Overall mean age was approximately 8 years and most children were classified as having bilateral spastic CP (75%) with about a third classified in GMFCS level III and another third in GMFCS level IV.

Creation of rGMFM-66
The creation of the rGMFM-66 is described in detail in Figure S2. In summary, we split the GMFM-66 data (n=1217 GMFM-66 measurements) into training data (for training and tuning the statistical models, approximately two-thirds of the whole data) and validation data (for validating the models). The GMFM-66 score was defined as
the result of the full GMFM-66 assessment, when evaluating all 66 GMFM-66 items. The training data were sorted by increasing GMFM-66 score. Then two points were defined that divided the ordered data into three parts, each containing a third of the training data (three terciles). The first tercile contained the measurements with the lowest GMFM-66 scores, the third tercile with the highest. One of the self-learning algorithms used to create models in this study was the random forest model. Using the random forest model, it is possible to evaluate the importance of the independent variable (here: individual GMFM-66 item) on the explained/dependent variable (here: GMFM-66 score). Thus, we created random forest models for each tercile to determine the least important GMFM-66 item for predicting the GMFM-66 score and calculated the mean squared error (MSE) of the model.

Then this GMFM-66 item was omitted in the creation of the next random forest models (now with 65 items) for each tercile and the MSE was again calculated. This procedure was repeated 50 times (meaning a reduction of the number of used GMFM-66 items from 66 to 16 for the prediction of GMFM-66 score). With this method, it was possible to evaluate the maximum number of omitted GMFM-66 items without an increase of MSE in the random forest model. Thus, there were three reduced GMFM-66 item sets, each one optimal for each tercile of the possible GMFM-66 score. Typically, it is not known in which tercile the GMFM-66 score will be before the assessment is performed. Therefore, the intersection of the three reduced GMFM-66 item sets was determined, which included seven items, and used to predict the tercile in which the GMFM-66 score would lie. Three self-learning algorithms (random forest, support vector machine [SVM], and feed forward neural net [FNN]) were evaluated for predicting the tercile. According to the prediction, the corresponding GMFM-66 items set can then be used.

In summary, in clinical application of the rGMFM-66, at first the same seven GMFM-66 items are tested and then the tercile predicted by a statistical model. According to the predicted tercile, the remaining GMFM-66 items of the corresponding tercile are tested as well. Then the GMFM-66 score is predicted using the measured GMFM-66 items.

### Calculation of the statistical models

In this study, three different self-learning algorithms (random forest, SVM, and FNN) were used to predict the GMFM-66 tercile and the GMFM-66 score. A detailed description of the different self-learning algorithms is beyond the scope of this paper; thus, in the following text, the creation of the statistical models is summarized.

The training data (n=852 GMFM-66 assessments) were used to train and tune the models. Each model type (random forest, SVM, and FNN) had specific hyperparameters with which the model could be adjusted to make the best prediction (Table S1, online supporting information). For this purpose, the training data were resampled in two parts. One part was used to train the model and the second part was used for the adjustment of the hyperparameters of the model, the so-called model tuning. For random forest models the out-of-bag samples were used for model tuning, and for the SVM and FNN a 10-fold cross-validation was performed. The adjusted hyperparameters for random forest models were the number of trees, number of used GMFM-66 items for each split, sample size of the used data for a tree, and the minimum node number of terminal nodes. The adjusted hyperparameters for the SVM model were epsilon (epsilon of the insensitive-loss function), cost, and gamma (both are parameters of the kernel function [radial basis]). For more details, see the package description of the R package e1071.15

In this study we used FNN models with one hidden layer. For regression (prediction of GMFM-66 score), we used the function neuralnet in neuralnet package with rprop+ algorithm. The adjusted hyperparameters of the FNN models were the number of nodes in the hidden layer and the threshold of the error function as stopping criteria.16 For classification (prediction of the tercile [first, second, or third]), we used mnet function in package mnet.

The adjusted hyperparameters of the FNN models were the number of nodes in the hidden layer and decay (a parameter for weight decay); for details see the package description.17

### Validation of the statistical models

To determine how well the rGMFM-66 agreed with the GMFM-66 score, each case in the validation sample was submitted to the full proposed assessment process:

### Table 1: Demographics of the study population, by sex and GMFCS level

| GMFCS level, % | I | II | III | IV | V |
|----------------|---|----|-----|----|---|
| Female         | 350 | 159 | 80  | 852 | 365 | 187 |
| Age, y:mo      | 8:10 | 8:9 | 8:5 | (3:9) | (3:10) | (3:10) |
| Height, Z-score | −1.57 (1.31) | −1.67 (1.22) | −1.59 (1.32) |
| BMI, Z-score    | −0.76 (1.54) | −0.79 (1.42) | −0.67 (1.40) |

| Data are mean (SD) unless otherwise indicated. GMFCS, Gross Motor Function Classification System; BMI, body mass index. |
selection of tercile by the seven items and scoring of the chosen item set. The ability of the three different models to predict the GMFM-66 score was evaluated with intra-class correlation coefficient (ICC) and Bland–Altman statistics by using the validation data (n=365, prediction of GMFM-66 score, single time). In addition, we analysed the ability of the best model in predicting the change over time in the study population with two complete GMFM-66 assessments 1 year apart (n=187). These results were compared with the diagnostic performance of the already existing GMFM-66-IS by Russell et al.6

The change over time of the GMFM-66 score is often used to evaluate the effectiveness of a therapeutic intervention. Thus some authors recommend comparing the observed change over time with minimal clinically important differences. For this reason, we evaluated the diagnostic performance of the rGMFM-66 to detect a relevant change in the GMFM-66 score. All children with GMFM-66 remeasurement after 12 months (n=187) were classified according to a predefined limit as having a change in GMFM-66 score or not. The same classification was performed with the rGMFM-66 and GMFM-66-IS to evaluate their diagnostic performances. The predefined limit was set to values between −10.0 and +10.0 in 0.1 steps.

Statistical analysis
All calculations were performed with RStudio version 1.3.959 in conjunction with R version 4.0.1 (R Foundation for Statistical Computing, Vienna, Austria) and the statistical packages neuralnet version 1.44.2, ranger version 1.12.1, randomForest version 4.6-14, e1071 version 1.7-3, rel 1.4.2, and BlandAltmanLeh version 0.3.1. The results are presented as mean (SD) or count (relative frequency), unless otherwise stated. Values of p<0.05 were set as the threshold for statistical significance. The ICCs were calculated with the specifications two-way model, agreement, and single measurement (function icc in the rel package). The p values for multiple comparisons (Fig. S3, online supporting information) were calculated with pairwise comparisons using Wilcoxon rank-sum tests with continuity correction and p values adjustment according to the Bonferroni–Holm method.

RESULTS
Study population
There were no significant differences in the demographics between training and validation data for single time and change over time analysis (Table 1).

Creation of the rGMFM-66
Figure S4 (online supporting information) gives the MSE of the random forest models with reducing numbers of used GMFM-66 items. To predict the GMFM-66 score in the first tercile (of the training data), the 41 least important GMFM-66 items could be omitted without a relevant increase in the MSE of the random forest model. For the second tercile 32 GMFM-66 items could be omitted without a relevant increase in the MSE, and for the third tercile 25 items (Fig. S4). This meant that to predict a GMFM-66 score, in the first tercile 25 GMFM items were needed, in the second tercile 34, and in the third tercile 41. Figure S5 (online supporting information) summarizes the GMFM-66 items used in each tercile.

Validation of the statistical models
To evaluate the prediction of the GMFM-66 tercile, the false classification rates of the three models (using the validation data) were calculated as 7.4% for the random forest model, 7.1% for the SVM model, and 6.6% for the FNN model. Thus, the FNN model with five nodes in the hidden layer and a threshold of 0.215 performed best in predicting the GMFM-66 items used in each tercile.

| Table 2: Criterion validity of the reduced GMFM-66 and GMFM-66-IS with the GMFM-66 |
| --- |
| Single time (n=378) |
| **Random forest** | **Bland–Altman statistics** |
| Random forest | **ICC** | Upper limit (97.5%) | Lower limit (2.5%) |
| Support vector machine | 0.995 (0.994–0.996) | 3.34 (3.07–3.68) | –3.33 (–3.63 to –3.02) |
| Feed forward neural net | 0.997 (0.996–0.997) | 2.90 (2.64–3.15) | –2.71 (–2.98 to –2.46) |
| GMFM-66-IS* | 0.996 (0.995–0.997) | 2.87 (2.60–3.13) | –2.95 (–3.21 to –2.68) |
| GMFM-66-IS* | 0.994 (0.993–0.996) | 3.80 (3.48–4.12) | –3.25 (–3.57 to –2.92) |
| **Change score (n=187)** |
| **Random forest** | **Bland–Altman statistics** |
| rGMFM-66 | **ICC** | Upper limit | Lower limit |
| rGMFM-66 with SVM | 0.993 (0.993–0.995) | 2.62 (2.29–2.96) | –2.60 (–2.93 to –2.26) |
| GMFM-66-IS* | 0.976 (0.968–0.982) | 4.74 (4.12–5.36) | –4.95 (–5.56 to –4.33) |

*The GMFM-66 item set consists of four different sets, of which set 1 has 15 items, set 2 has 29 items, set 3 has 39 items, and set 4 has 22 items. On average, 30.5 items were evaluated in the study population. GMFM-66, 66-Item Gross Motor Function Measure; GMFM-66-IS, 66-Item Gross Motor Function Measure item set; ICC, intraclass correlation coefficient; CI, confidence interval.
(random forest, SVM, and FNN) had higher ICCs than the GMFM-66-IS (Table 2). The SVM model had the highest ICC and the best Bland–Altman statistics of the evaluated models (Fig. 1). In the study population, 34.5 GMFM-66 items on average were used to calculate the rGMFM-66 score. The mean absolute error (MAE) of the prediction of the GMFM-66 score did not differ significantly between females and males ($p=0.118$) or CP subtypes ($p=0.131$).

There was a slight, but significant, difference in the MAE between the five GMFCS levels ($p<0.001$), with the lowest MAE for GMFCS levels III and IV. The median MAE for GMFCS levels III and IV was 0.7 to 0.8 and for GMFCS levels I, II, and V 1.1 to 1.6 points (Fig. S3).

The prediction of the GMFM-66 change score over 1 year was performed only with the SVM model (because it was the best model in the single-time validation). Table 2 also gives the results compared with the GMFM-66-IS. Table S2 (online supporting information) shows the results for the diagnostic performance of the rGMFM-66 compared with the GMFM-66-IS to detect a change in the full GMFM-66.

**DISCUSSION**

The study results indicate that by using self-learning algorithms it is possible to increase the efficiency of the full GMFM-66 assessment. The presented rGMFM-66 score showed a very good agreement with the GMFM-66 score when applied to a single assessment, as indicated by the ICC and Bland–Altman statistics (Table 2). The mean of the Bland–Altman statistics was almost 0, indicating that in our study population the rGMFM-66 was systematically neither higher nor lower than the full GMFM-66 score (Fig. 1). In addition, the 95% confidence interval (CI) of the Bland–Altman statistics was lower than ±3.0 points and the MAE was slightly over 1 point, indicating a very

---

**Figure 1**: (a–d) Bland–Altman diagrams for single time and change score validation of the reduced version of the 66-item Gross Motor Function Measure (rGMFM-66) and the item-set method of the GMFM-66 (GMFM-66-IS) against the GMFM-66. Dashed lines, 95% confidence intervals (CI) of the Bland–Altman diagrams. The widths of the 95% CI are an indication of the degree of agreement between the rGMFM-66 and GMFM-66, and GMFM-66-IS and GMFM-66. The rGMFM-66 shows a higher agreement than the GMFM-66-IS. CS, change score.
good score agreement in most of the study population (Table 2 and Fig. S3). The score agreement was independent of sex and CP subtype. Although the same items were always evaluated, there was also a good performance of the rGMFM-66 even with unilateral CP, probably because the rGMFM-66 contains almost no one-sided, 'asymmetrical' item (with exception of item 27: sit on mat: touches toy placed 45° behind child's left side; and item 66: stand, two hands on large bench: cruises five steps to left). There was slightly better score agreement for children classified in GMFCS levels III and IV than in the other GMFCS levels (Fig. S3). The reason for this could be that this group made up most of the study population.

When the change over time was measured, the rGMFM-66 score again showed very good agreement with the GMF66 score, with comparable ICC and Bland–Altman statistics as for the single-time evaluation (Table 2 and Fig. 1). Additionally, the diagnostic performance of the rGMFM-66 in detecting a relevant change in the GMF66 was excellent, as indicated by the very high sensitivity, specificity, positive predictive value, and negative predictive value (92.3–98.1%; Table S2).

**GMFM-66-IS**

Other abbreviated versions of the GMFM-66 were already introduced by Russell et al. (GMFM-66-IS)⁶ and by Brunton et al. (GMFM-66 B&C).⁷ Unfortunately, the values for the GMFM66 B&C were not available to us and therefore could not be compared with the rGMFM-66. However, Avery et al. were able to show that the GMFM-66-IS matches the GMFM-66 better than the GMFM-66 B&C.⁹ They reported the equal ICC for single-time evaluation, as in our study population (ICC 0.944 [95% CI 0.993–0.996]). The ICC for the change over time (1y) was slightly higher in our study population than that reported by Avery et al. (0.976 [95% CI 0.968–0.982] vs 0.922 [95% CI 0.888–0.946]).

**Comparison of rGMFM-66 and GMFM-66-IS**

The diagnostic performance of the rGMFM-66 in predicting the GMFM-66 score at a single time was significantly higher than that of the GMFM-66-IS, as indicated by the higher ICC and the lower Bland–Altman statistics (Table 2).

The time savings when performing an abbreviated version of the full GMFM-66 measurement are particularly important in clinical use. Since in the clinical setting the GMFM-66 is also very often used as an outcome parameter, the change over time is of particular importance. In this category too, the rGMFM-66 shows a better match with the GMFM-66 than the GMFM-66-IS (Table 2 and Table S2). In our study population, the mean number of measured GMFM-66 items was slightly higher in rGMFM-66 assessments than in GMFM-66-IS (34.5 vs 30.5 items).

Overall, the rGMFM-66 appeared to correspond more closely to the GMFM-66 than the GMFM-66-IS and seems to be a good alternative to the GMFM-66, when time saving is of particular interest. Additionally, since testing 66 motor items could be very exhausting for the child, depending on the condition, the rGMFM-66 may provide fewer confounding results.

**Study limitations**

Limitations of this study include selection bias because our artificial intelligence-powered algorithms were based on the data of our rehabilitation programme. Therefore, relatively few children with CP in GMFCS levels I and V were represented in this cohort. Also, the study population consisted largely of children with spastic CP. Since most children with CP have this CP subtype, these findings should be widely applicable to the population of children with CP.

In addition, our study was a retrospective study. Although we divided the existing data set into training, tuning, and validation sets for the validity check, we also considered it necessary to check the rGMFM-66 prospectively for the validity in other data sets. Another important limitation is that this study was undertaken using information from an existing data set, and a non-shortened version of the GMFM-66 was actually administered. On the other hand, this approach eliminated possible confounding factors for variation, for example fatigue of the children, differences among raters, or influence of the testing environment.

This study should show the feasibility of increasing the efficiency of the administration of the GMFM-66 by using self-learning algorithms. The presented results cannot currently be easily used in clinical practice. For clinical use, the results would first have to be implemented in software, such as the Gross Motor Ability Estimator (version 2) scoring software for the GMFM-66.

**CONCLUSION**

The study results indicate that by using self-learning algorithms it was possible to increase the efficiency of administration of the GMFM-66 assessment. The presented rGMFM-66 score showed an excellent agreement with the GMFM-66 score when applied to a single assessment and when evaluating the change over time.

**ACKNOWLEDGEMENTS**

We thank the physiotherapists of the UniReha GmbH, Centre of Prevention and Rehabilitation, programme ‘Auf die Beine’ (University of Cologne) for their dedication to their work and assessment of the GMFM-66. We thank Valerie Adames for GMFM-66 quality management. We thank Angelika Stabrey for data handling and Ida Alperstedt for data input. We thank all patients and families for participating in the Cologne physiotherapy treatment programme ‘Auf die Beine’. Novotec Medical GmbH is supporting research at UniReha GmbH. Christina Stark received travel support from Novotec Medical GmbH to present research at medical conferences. She received research grants by Novotec Medical GmbH. She has no further financial relationships with Novotec Medical GmbH. No other disclosures are reported. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Reduced GMFM-66 in Cerebral Palsy Ibrahim Duran et al. 233
DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy, ethical restrictions and potentially commercialization of research findings.

SUPPORTING INFORMATION
The following additional material may be found online:

Figure S1: Consort diagram of the study.

REFERENCES
1. Damiano DL. Meaningfulness of mean group results for determining the optimal motor rehabilitation program for an individual child with cerebral palsy. Dev Med Child Neurol 2014; 56: 1141-6.
2. Rosenbaum PL, Walter SD, Hanna SE, et al. Prognosis for gross motor function in cerebral palsy: creation of motor development curves. JAMA 2002; 288: 1357-63.
3. Russell DJ, Avery LM, Rosenbaum PL, Raina PS, Walter SD, Palisano RJ. Improved scaling of the Gross Motor Function Measure for children with cerebral palsy: evidence of reliability and validity. Phys Ther 2000; 80: 873–85.
4. Avery LM, Russell DJ, Raina PS, Walter SD, Rosenbaum PL. Rasch analysis of the gross motor function measure: validating the assumptions of the Rasch model to create an interval-level measure. Arch Phys Med Rehabil 2003; 84: 697-705.
5. Russell DJ, Rosenbaum P, Wright M, Avery LM. Gross Motor Function Measure (GMFM-66 & GMFM-88) User’s Manual. 2nd edn. London, UK: Mac Keith Press, 2013.
6. Russell DJ, Avery LM, Walter SD, et al. Development and validation of item sets to improve efficiency of administration of the 66-item Gross Motor Function Measure in children with cerebral palsy. Dev Med Child Neurol 2010; 52: e48-54.
7. Brunton LK, Bartlett DJ. Validity and reliability of two abbreviated versions of the Gross Motor Function Measure. Phys Ther 2011; 91: 577–88.
8. Briganti G, Le Moine O. Artificial intelligence in medicine: today and tomorrow. Front Med 2020; 7: 27.
9. Patel VL, Shortliffe EH, Stefaneli M, et al. The coming of age of artificial intelligence in medicine. Artif Intell Med 2009; 46: 5–17.
10. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol 2019; 19: 66.
11. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019; 20: e262–73.
12. Saeys Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics. Bioinformatics 2007; 23: 2507–17.
13. Duran I, Stark C, Martakis K, Hamacher S, Semler O, Schoenau E. Reference centiles for the gross motor function measure and identification of therapeutic effects in children with cerebral palsy. J Eval Clin Pract 2019; 25: 78-87.
14. Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. Dev Med Child Neurol 1997; 39: 214–23.
15. Meyer D, Dimitriadou E, Hornik AK. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7-3. https://CRAN.R-project.org/package=e1071 (accessed 18 March 2021).
16. Fritsch S, Guenther F, Wright MN. neuralnet: Training of Neural Networks: R package version 1.44.2. https://CRAN.R-project.org/package=neuralnet (accessed 18 March 2021).
17. Venables WN, Ripley BD. Modern Applied Statistics with S. 4th edn. New York, NY: Springer, 2007.
18. Oeffinger D, Bagley A, Rogers S, et al. Outcome tools used for ambulatory children with cerebral palsy: responsiveness and minimum clinically important differences. Dev Med Child Neurol 2008; 50: 918–25.
19. Avery LM, Russell DJ, Rosenbaum PL. Criterion validity of the GMFM-66 item set and the GMFM-66 basal and ceiling approaches for estimating GMFM-66 scores. Dev Med Child Neurol 2013; 55: 534–8.