An Exploratory Study on Techniques for Quantitative Assessment of Stroke Rehabilitation Exercises

Min Hun Lee
Daniel P. Siewiorek
Asim Smailagic
Carnegie Mellon University
{minhunl,dps,asim}@cs.cmu.edu

Alexandre Bernardino
Instituto Superior Técnico
alex@isr.tecnico.ulisboa.pt

Sergi Bermúdez i Badia
Madeira Interactive Technology Institute, University of Madeira, NOVA-LINCS
sergi.bermudez@m-iti.org

ABSTRACT

Technology-assisted systems to monitor and assess rehabilitation exercises have an opportunity of enhancing rehabilitation practices by automatically collecting patient’s quantitative performance data. However, even if a complex algorithm (e.g. Neural Network) is applied, it is still challenging to develop such a system due to patients with various physical conditions. The system with a complex algorithm is limited to be a black-box system that cannot provide explanations on its predictions. To address these challenges, this paper presents a hybrid model that integrates a machine learning (ML) model with a rule-based (RB) model as an explainable artificial intelligence (AI) technique for quantitative assessment of stroke rehabilitation exercises. For evaluation, we collected therapist’s knowledge on assessment as 15 rules from interviews with therapists and the dataset of three upper-limb stroke rehabilitation exercises from 15 post-stroke and 11 healthy subjects using a Kinect sensor. Experimental results show that a hybrid model can achieve comparable performance with a ML model using Neural Network, but also provide explanations on a model prediction with a RB model. The results indicate the potential of a hybrid model as an explainable AI technique to support the interpretation of a model and fine-tune a model with user-specific rules for personalization.

CCS CONCEPTS

• **Human-centered computing → Interactive systems and tools**; • **Applied computing → Health care information systems**.

KEYWORDS

Human-AI Interaction; Explainable AI; Decision Support Systems; Human Activity Recognition; Stroke Rehabilitation Assessment

ACM Reference Format:
Min Hun Lee, Daniel P. Siewiorek, Asim Smailagic, Alexandre Bernardino, and Sergi Bermúdez i Badia. 2020. An Exploratory Study on Techniques for Quantitative Assessment of Stroke Rehabilitation Exercises. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization (UMAP ’20), July 14–17, 2020, Genoa, Italy*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3340631.3394872

ACM ISBN 978-1-4503-6861-2/20/07…$15.00
https://doi.org/10.1145/3340631.3394872

1 INTRODUCTION

Supervised physical rehabilitation sessions are one of effective ways to improve the functional ability of patients with musculoskeletal and neurological disorders (e.g. stroke) [18]. However, as therapists have limited availability, they often prescribe in-home rehabilitation, where patients perform exercises themselves [3]. Therapists mainly rely on a patient’s self-report to discuss patient’s process and adjust a treatment intervention [18]. Without any quantitative performance data, therapists encounter difficulty with making informed decision on patient’s intervention.

With recent improvement on sensor and machine learning (ML) techniques, researchers have demonstrated the feasibility to develop a technology-assisted rehabilitation monitoring system that can provide objective kinematic analysis on patient’s functional status to support therapist’s decision making and improve rehabilitation practices [17, 25]. Most related work on human activity recognition and rehabilitation assessment focuses on improving the performance of a ML model by applying complex algorithms [9, 15, 19]. However, it is challenging to derive a model that can replicate therapist’s monitoring and assessment due to various conditions of patients. In addition, as a model with complex algorithms cannot explain its prediction to support therapist’s decision making, therapists can lose trust on it and abandon its usage [13, 17].

This paper presents a hybrid model that integrates a data-driven, machine learning (ML) model with a rule-based (RB) model using a weighted average ensemble technique [2, 14, 16] to assess the quality of motion. For the development, we conducted a semi-structured interview with therapists to elicit their knowledge of assessing rehabilitation exercises into 15 rules and collected the dataset of three upper-limb exercises from 15 post-stroke and 11 healthy subjects and the annotations of the dataset from therapists. Our experimental results demonstrate that a hybrid model outperforms a rule-based model, and achieves comparable performance with a data-driven ML model using Neural Network while providing a possibility of interpreting a model by analyzing rules of the RB model to fine-tune a model for personalization.

Although prior work demonstrated the feasibility of monitoring and assessing rehabilitation exercises [15, 25], it is still challenging to develop a system that can be utilized by therapists [13, 17]. As a first step to support deployment of such a system in practice, this paper presents a hybrid model as an explainable artificial intelligence (AI) technique and compare it with two widely applied techniques: data-driven machine learning (ML) and rule-based (RB) models. This work aims to broaden knowledge on an explainable AI technique for human activity recognition and understanding (e.g. quantitative assessment of stroke rehabilitation exercises).

2 RELATED WORK

Researchers have investigated the possibility of automatically monitoring and assessing chronic diseases with computational models [15, 25] to provide patients feedback without the presence of a therapist and supplement therapist's decision making on patient's treatment with quantitative data [17]. The approaches of these computational models can be categorized into either a rule-based (RB) or machine learning (ML) model.

A rule-based (RB) model requires the engagement of domain experts to derive a set of monitoring rules [14]. For example, Huang conducted an experiment on whether an authoring tool can support therapists to specify repetitions and joint angles for monitoring knee rehabilitation exercises [12]. This rule-based approach can be easily modularized and flexibly recombined to develop a customized monitoring model. However, it is time consuming to manually review a large amount of sensor measurements and derive a set of rules to monitor the status of an individual. In addition, it is difficult to articulate therapist's decision making on complex and abstract concepts into a set of rules.

Alternative approach is to utilize a machine learning algorithm with labeled sensor data [14]. Although this approach has the benefit of automatically learn a meaningful function (e.g. Neural Networks) to assess the quality of motion [5, 15], it is challenging to replicate therapist's evaluation due to patients with various physical conditions. In addition, this approach with a complex algorithm cannot provide explanations on its predictions to support therapist's decision making [10], which can exacerbate therapist's user experience and impede its adoption in practices [13, 17].

Explainability has been an actively explored by researchers to create a better machine learning model with improved transparency and user acceptance [4, 6, 7]. However, it is still challenging to get full interpretation on how a complex model works [1]. Holzinger describes the necessity of constructing contextual explanatory models on real-world phenomena for an explainable artifical intelligence (AI) [11]. As a rule-based (RB) approach has the benefit of being comprehensible, this paper hypothesizes that such a RB model can serve as a contextual explanatory model that can supplement a complex machine learning (ML) model. This paper derives a hybrid approach that integrates a ML model with an interpretable RB model to increase the interpretability of a model. This work contributes to increase knowledge on an explainable artificial intelligence (AI) technique for human activity recognition and understanding (e.g. quantitative assessment of stroke rehabilitation exercises).

3 SPECIFICATIONS OF THE STUDY FOR STROKE REHABILITATION

We selected a probe domain as stroke, which is the second leading cause of death and third most common contributor to disability [8]. We had iterative discussion with three therapists (μ = 6.33, σ = 2.05 years of experience in stroke rehabilitation) in Table 1 and specified the designs of our study on stroke rehabilitation: three upper-limb exercises and performance components for assessment [17].

3.1 Three Task-Oriented Upper Limb Exercises

Three upper-limb stroke rehabilitation exercises (Figure 1) are recommended by therapists due to their correspondence with major motion patterns [17]: elbow flexion for Exercise 1, shoulder flexion for Exercise 2, elbow extension for Exercise 3. For Exercise 1, a subject has to raise subject's wrist to the mouth as if drinking water with a cup. For Exercise 2, a subject has to pretend touching a light switch on the wall. For Exercise 3, a subject has to practice the usage of a cane by extending subject's elbow in the seated position.

Table 1: The participants of the specification, of the annotation, of the rule elicitation (ElicitRule)

ID		Studies		# of Years in
	Specification	Annotation	ElicitRule	Stroke Rehab
TP1	✓	✓	✓	6
TP2	✓	✓	✓	4
TP3	✓			9

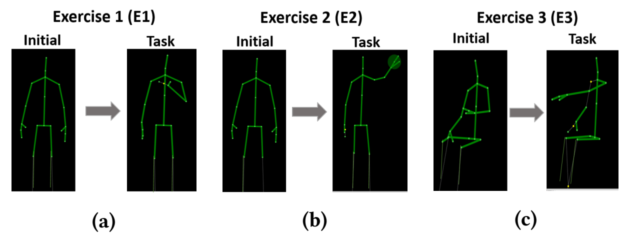

Figure 1: (a) Exercise 1 (E1): *'Bring a Cup to the Mouth'* (b) Exercise 2 (E2): *'Switch a Light On'* (c) Exercise 3 (E3): *'Move a Cane Forward'*

3.2 Performance Components

Three common performance components are identified to assess the quality of motion [17]: *'Range of Motion (ROM)'*, *'Smoothness'*, and *'Compensation'*, which are based on commonly used stroke assessment tools (i.e. Fugl Meyer Assessment [23] and Wolf Motor Function Test [24]). The *'ROM'* indicates how closely a patient performs a task-oriented exercise. The *'Smoothness'* checks the degree of trembling and irregular movement of joints while performing an exercise. The *'Compensation'* monitors whether a patient performs any compensated movements to achieve a target movement. For instance, a patient might elevate his/her shoulder to raise the affected hand [17]. Each patient might have different compensated movements based on patient's functional status: one patient might elevate shoulder while rotating trunk and the other patient might elevate shoulder while leaning trunk backward [17]. We denote the correct/normal performance component as $Y = 1$ and incorrect/abnormal performance component as $Y = 0$.

3.3 Kinematic Features

We represent an exercise motion with sequential joint coordinates from a Kinect v2 sensor (Microsoft, Redmond, USA) and extract various kinematic features [15].

For the *'ROM'*, we compute joint angles (e.g. elbow flexion, shoulder flexion, elbow extension) and normalized relative trajectory (i.e. Euclidean distance between two joints - head and wrist, head and elbow). For the *'Smoothness'*, we compute the following speed related features: speed, acceleration, jerk, zero crossing ratio of acceleration and jerk, and Mean Arrest Period Ratio (the portion of the frames when speed exceeds 10% of the maximum speed) [22]. As we have upper-limb exercises, we computed these speed related features on wrist and elbow joints. For the *'Compensation'*, we compute joint angles (i.e. the elevated angle of shoulder, the tilted angle of spine, and shoulder abduction) and normalized trajectories (distances between joint positions of head, spine, shoulder in x, y, z axis from the initial to current frame).

A moving average filter with the window size of five frames is applied to reduce noise of acquiring joint positions from a sensor similar to [15]. Given an exercise motion, we compute a feature matrix ($\mathbf{F} \in R^{t \times d}$) with t frame and d features and statistics (i.e. max, min, range, average, and standard deviation) over all frames of the exercise to summarize a motion into a feature vector ($X \in R^{5d}$).

4 QUANTITATIVE ASSESSMENT OF REHABILITATION EXERCISES

4.1 Machine Learning (ML) Model

A machine learning (ML) model utilizes a supervised learning algorithm to predict the quality of motion or compute the posterior probability of being normal/correct, $P_{PM} = P(Y = 1|X)$ on each performance component, where X refers to the feature vector of an exercise motion and $Y \in \{0, 1\}$ describes the correctness on a performance component of a motion. We explore various traditional algorithms [15]: Decision Tree (DT), Linear Regression (LR), Support Vector Machine (SVM), Neural Network (NN) using the *'Scikit-learn'* [21] and the *'PyTorch'* libraries [20].

For DT, Classification and Regression Trees (CART) is utilized to build prune trees. For LR, we apply $L1, L2$ regularization or linear combination of $L1$ and $L2$ (ElasticNet with 0.5 ratio) to avoid overfitting. For SVM, we apply either linear or Radial Basis Function (RBF) kernels with penalty parameter, $C = 1.0$. NN is trained while grid-searching over various architectures (i.e. one to three layers with $32, 64, 128, 256, 512$ hidden units) and different learning rates (i.e. $0.0001, 0.005, 0.001, 0.01, 0.1$). NN applies the *'ReLu'* activation functions and *'AdamOptimizer'* and is trained until the tolerance of optimization is 0.0001 or 200 iterations.

4.2 Rule-based (RB) Model

A rule-based (RB) model utilizes a set of feature-based rules from therapists to estimate the quality of motion. For the development, we conducted a semi-structured interview with two therapists (Table 1) to elicit their knowledge of assessing stroke rehabilitation exercises, which is formalized as 15 independent *if-then* rules. For example, the rule of the *'ROM'* for Exercise 1 is specified as follows:

$$\hat{Y} = \begin{cases} 1 & \text{if } p^{max}(wr, c_y) >= p^{max}(spsh, c_y) \\ 0 & \text{else} \end{cases}$$

where $p(j, c)$ indicates a joint position with a joint j (e.g. wrist (wr) and spine shoulder, the top of spine, ($spsh$)) and the coordinate of a joint, c in the set $C \in \{c_x, c_y, c_z\}$. $\hat{Y}$ denotes the predicted label on a performance component. This rule compares the maximum position of wrist joint, $p^{max}(wr, c_y)$ with that of spine shoulder joint, $p^{max}(spsh, c_y)$ in the y-coordinate to roughly estimate whether a patient achieves a target position of Exercise 1 (Figure 1a).

A rule-based (RB) model computes a score of being correct on performance component as follows:

$$P_{RB} = \frac{1}{|\mathbb{R}|} \sum_{r \in \mathbb{R}} \min(\frac{f_r}{\tau_r}, 1) \tag{1}$$

where f_r indicates the feature value of a rule r from a exercise motion (e.g. $p^{max}(wr, c_y)$ for the example above), τ_r describes the threshold value of a rule r (e.g. $p^{max}(spsh, c_y)$ for the example above). $\mathbb{R}$ describes the set of rules from the therapists. min function is applied so that this equation assigns a value of 1 if the feature value of a rule exceeds the threshold of that rule. Otherwise, the equation normalizes the feature value of a rule with the threshold of a rule to compute the likelihood of being correct.

4.3 Hybrid Model

A hybrid model (HM) applies a weighted average, ensemble technique [2, 14, 16] to combine two perspectives on assessment : a data-driven, machine learning (ML) model and a rule-based (RB) model from therapists. The performance of each model (i.e. F1-score in $[0, 1]$) is utilized as the weight of each model. A hybrid model computes the score of being correct, $P_{HM} = P(Y = 1|X)$ as follows:

$$P_{HM} = \frac{\rho_{ml}}{\rho_{ml} + \rho_{rb}} P_{ML} + \frac{\rho_{rb}}{\rho_{ml} + \rho_{rb}} P_{RB} \tag{2}$$

where P_{ML} and P_{RB} indicate the scores of ML and RB models and ρ_{ml} and ρ_{rb} describe F1-scores of ML and RB models respectively.

5 DATASET OF THREE EXERCISES

The dataset of three exercises is collected from 15 post-stroke and 11 healthy subjects using a Kinect v2 sensor [15]. During the data collection, a sensor was located at the height of 0.72m above the floor and 2.5m away from a subject and recorded trajectory of joints and video frames at 30 Hz. The starting and ending frames of exercise movements were manually annotated.

After signing the consent form, a subject participated in the data collection. Fifteen post-stroke patients (2 females) with diverse functional abilities from mild to severe impairment (37 ± 21 out of 66 Fugl Meyer Scores [23]) performed 10 repetitions of each exercise with both affected and unaffected sides. Eleven healthy subjects (1 female) performed 15 repetitions of each exercise with their dominant sides.

Two therapists (TP 1 and TP 2 in Table 1) annotated the dataset to implement our approach and compute therapist's agreement level. They individually watched the recorded videos of patient's exercise motions and annotated the performance components of exercise motions in the dataset. For evaluation, we utilize the annotation of therapist 1 (TP 1), who evaluated patient's functional ability with Fugl Meyer Assessment, as the ground truth. The annotation of therapist 2 (TP 2) is compared with that of TP1 to measure therapist's agreement using F1-scores (TPA in Table 2).

6 RESULTS

The implementation of models is evaluated with leave-one-subject-out cross validation on post-stroke patients, which trains a model with data from all subjects except one testing subject and tests a model with affected motions of the left-out post-stroke subject.

Table 2 summarizes the performance of models, which measures an agreement with the annotation of therapist 1 (TP 1) by computing average F1-scores on performance components of three exercises. The parameters of Neural Networks (i.e. hidden layers/units and learning rate) that achieve the best F1-score during leave-one-subject-out cross validation are summarized in Table 3.

Table 2: Performance (F1-scores) of machine learning (ML) models with various algorithms, rule-based (RB) models, hybrid models (HM), and therapist's agreement (TPA)

	Exercise 1 (E1)	Exercise 2 (E2)	Exercise 3 (E3)	Overall
ML - DT	0.6901 ± 0.0405	0.7645 ± 0.0867	0.6488 ± 0.0412	0.7011 ± 0.0769
ML - LR	0.7246 ± 0.0593	0.6430 ± 0.0982	0.7267 ± 0.0391	0.6981 ± 0.0801
ML - SVM	0.7232 ± 0.0364	0.6971 ± 0.0891	0.7410 ± 0.0052	0.7204 ± 0.0585
ML - NN	**0.8632 ± 0.0816**	**0.8388 ± 0.0518**	**0.7818 ± 0.0096**	**0.8279 ± 0.0605**
RB	0.6148 ± 0.1702	0.6932 ± 0.1630	0.4384 ± 0.1569	0.5821 ± 0.1066
HM	0.8437 ± 0.0697	0.7545 ± 0.0561	0.7812 ± 0.0479	0.7931 ± 0.06440
TPA	0.8120 ± 0.1458	0.7790 ± 0.1324	0.7654 ± 0.1382	0.7854 ± 0.0195

Table 3: Parameters of Neural Networks

	Hidden Layers and Units / Learning Rate		
	ROM	Smooth	Comp
E1	(32, 32, 32) / 0.1	(16) / 0.0001	(256, 256) / 0.1
E2	(256) / 0.1	(512, 512) / 0.1	(128) / 0.1
E3	(256) / 0.1	(64, 64) / 0.001	(128, 128) / 0.1

For the machine learning (ML) models, the performance of a model is inversely proportional to the interpretability of a model similar to [10]. ML models with complex algorithms (i.e. Support Vector Machine, ML-SVM or Neural Network, ML-NN) perform better than ML models with interpretable algorithms (e.g. Decision Trees, ML-DT or Linear Regression, ML-LR) while sacrificing the interpretability on a model. Specifically, ML models with Neural Networks (ML-NN) outperform ML models with other algorithms: Decision Trees, ML-DT (0.7011 average F1-scores), Linear Regression, ML-LR (0.6981 average F1-scores), Support Vector Machine, ML-SVM (0.7204 average F1-scores). ML-NN achieves a good agreement level with therapist 1 (TP 1)'s annotation (i.e. 0.8279 average F1-scores over three exercises), which is 0.04 higher average F1-score than therapist's agreement (TPA) between TP 1 and TP2 (i.e. 0.7854 average F1-scores over three exercises).

In contrast, the rule-based (RB) model achieves the lowest agreement level with therapist 1's annotation: 0.5821 average F1-scores over three exercises. According to further analysis on the RB model, we found that such low performance occurred, because elicited rules from therapists are generic and not tuned for individuals with different physical conditions. For instance, one rule of monitoring the *'Smoothness'* is to check whether the zero-crossing ratio of a wrist acceleration (i.e. the period of a motion, in which a sign of acceleration changes) on the y-axis exceeds 20% or not. However, we observed that some smoothly coordinated motions from patients have 25 - 35 % ratio on this feature and mis-classified as *'Incorrect'*. This implies the necessity of generating personalized rules for patients with various physical characteristics and functional abilities. Thus, we expect if a system can present patient's feature values on rules to therapists, therapists might be able to tune these generic rules for personalized rehabilitation assessment [16].

The hybrid model (HM) combines the machine learning model with Neural Networks (ML-NN) and the rule-based (RB) model and achieves 0.7931 average F1-scores over three exercises. The performance of the HM is 0.03 lower average F1-scores than that of the machine learning model with with Neural Network (ML-NN). Although combining two perspectives on assessment (i.e. ML-NN and RB models) does not improve the performance of a model due to the RB models with generic rules, the HM still achieves higher performance than ML models with other algorithms (i.e. Decision Tree, Linear Regression, Support Vector Machine). Compared to ML-NN, the HM has a potential benefit of interpreting a model by analyzing rules of the RB model and fine-tuning a model with patient-specific rules [16]. In addition, the HM achieves good agreement with therapist 1's annotation, which is equally good with therapist's agreement (TPA) between TP 1 and TP 2. The HM shows a potential to consistently replicate therapist's assessment.

7 CONCLUSION

In this paper, we present a hybrid model that integrates a data-driven, machine learning (ML) model with an interpretable rule-based (RB) model from therapists as an explainable artificial intelligence (AI) technique for quantitative assessment of stroke rehabilitation exercises, and compare it with two widely used approaches of prior work (i.e. ML and RB models). Our results show that a hybrid model can achieve good performance that is comparable to the performance of the ML model with Neural Network (ML-NN), but also provides an opportunity of interpreting a model by analyzing rules of the RB model and tuning a generic rule for personalized rehabilitation assessment. Yet, a further study is necessary to investigate the feasibility of presenting features and tuning a model.

In contrast to most related work on modeling and understanding human activities (e.g. exercises) that focuses on improving the performance of a model by applying a complex deep learning model [9, 15, 19], our work highlights the importance on creating an explainable model to support the deployment of a model in practice. We believe this study on a hybrid model can be a valuable reference on an explainable artificial intelligence (AI) technique for human activity recognition and understanding (e.g. quantitative assessment of rehabilitation exercises).

ACKNOWLEDGMENTS

The authors appreciate all subjects and therapists for their participation in the study. This work is partially supported by the FCT (SFRH/BD/113694/2015, LARSyS - Plurianual funding 2020-2023), the CMU-PT IntelligentCare project (LISBOA-01-0247-FEDER-045948), and the National Science Foundation (NSF) under grant number CNS-1518865.

REFERENCES

[1] Amina Adadi and Mohammed Berrada. 2018. Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 6 (2018), 52138–52160.

[2] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. 2019. Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41, 2 (2019), 423–443.

[3] Mark T Bayley, Amanda Hurdowar, Carol L Richards, Nicol Korner-Bitensky, Sharon Wood-Dauphinee, Janice J Eng, Marilyn McKay-Lyons, Edward Harrison, Robert Teasell, Margaret Harrison, et al. 2012. Barriers to implementation of stroke rehabilitation evidence: findings from a multi-site pilot project. *Disability and rehabilitation* 34, 19 (2012), 1633–1638.

[4] Or Biran and Courtenay Cotton. 2017. Explanation and justification in machine learning: A survey. In *IJCAI-17 workshop on explainable AI (XAI)*, Vol. 8. 1.

[5] Samarjit Das, Laura Trutoiu, Akihiko Murai, Dunbar Alcindor, Michael Oh, Fernando De la Torre, and Jessica Hodgins. 2011. Quantitative measurement of motor symptoms in Parkinson's disease: A study with full-body motion capture data. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 6789–6792.

[6] Finale Doshi-Velez and Been Kim. 2017. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608* (2017).

[7] Mengnan Du, Ninghao Liu, and Xia Hu. 2018. Techniques for interpretable machine learning. *arXiv preprint arXiv:1808.00033* (2018).

[8] Valery L Feigin, Bo Norrving, and George A Mensah. 2017. Global burden of stroke. *Circulation research* 120, 3 (2017), 439–448.

[9] Yu Guan and Thomas Plötz. 2017. Ensembles of Deep LSTM Learners for Activity Recognition Using Wearables. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 2, Article 11 (June 2017), 28 pages. https://doi.org/10.1145/3090076

[10] David Gunning. 2017. Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA)* 2 (2017).

[11] Andreas Holzinger. 2018. From machine learning to explainable AI. In *2018 World Symposium on Digital Intelligence for Systems and Machines (DISA)*. IEEE, 55–66.

[12] Kevin Huang. 2015. *Exploring in-home monitoring of rehabilitation and creating an authoring tool for physical therapists*. Ph.D. Dissertation. Carnegie Mellon University.

[13] Saif Khairat, David Marc, William Crosby, and Ali Al Sanousi. 2018. Reasons for physicians not adopting clinical decision support systems: critical analysis. *JMIR medical informatics* 6, 2 (2018), e24.

[14] Min Hun Lee. 2019. An Intelligent Decision Support System for Stroke Rehabilitation Assessment. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '19)*. ACM, 694–696. https://doi.org/10.1145/3308561.3356106

[15] Min Hun Lee, Daniel P. Siewiorek, Asim Smailagic, Alexandre Bernardino, and Sergi Bermúdez i Badia. 2019. Learning to Assess the Quality of Stroke Rehabilitation Exercises. In *Proceedings of the 24th International Conference on Intelligent User Interfaces (IUI '19)*. ACM, 218–228. https://doi.org/10.1145/3301275.3302273

[16] Min Hun Lee, Daniel P. Siewiorek, Asim Smailagic, Alexandre Bernardino, and Sergi Bermúdez i Badia. 2020. Interactive Hybrid Approach to Combine Machine and Human Intelligence for Personalized Rehabilitation Assessment. In *Proceedings of the ACM Conference on Health, Inference, and Learning (CHIL '20)*. ACM, 160–169. https://doi.org/10.1145/3368555.3384452

[17] Min Hun Lee, Daniel P. Siewiorek, Asim Smailagic, Alexandre Bernardino, and Sergi Bermúdez i Badia. 2020. Opportunities of a Machine Learning-based Decision Support System for Stroke Rehabilitation Assessment. arXiv:2002.12261

[18] Susan B O'Sullivan, Thomas J Schmitz, and George Fulk. 2019. *Physical rehabilitation*. FA Davis.

[19] Madhuri Panwar, Dwaipayan Biswas, Harsh Bajaj, Michael Jöbges, Ruth Turk, Koushik Maharatna, and Amit Acharyya. 2019. Rehab-Net: Deep Learning Framework for Arm Movement Classification Using Wearable Sensors for Stroke Rehabilitation. *IEEE Transactions on Biomedical Engineering* 66, 11 (2019), 3026–3037.

[20] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in PyTorch. (2017).

[21] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *Journal of machine learning research* 12, Oct (2011), 2825–2830.

[22] Brandon Rohrer, Susan Fasoli, Hermano Igo Krebs, Richard Hughes, Bruce Volpe, Walter R Frontera, Joel Stein, and Neville Hogan. 2002. Movement smoothness changes during stroke recovery. *Journal of Neuroscience* 22, 18 (2002), 8297–8304.

[23] Julie Sanford, Julie Moreland, Laurie R Swanson, Paul W Stratford, and Carolyn Gowland. 1993. Reliability of the Fugl-Meyer assessment for testing motor performance in patients following stroke. *Physical therapy* 73, 7 (1993), 447–454.

[24] Edward Taub, David M Morris, Jean Crago, Danna Kay King, Mary Bowman, Camille Bryson, Staci Bishop, Sonya Pearson, and Sharon E Shaw. 2011. Wolf motor function test (WMFT) manual. *Birmingham: University of Alabama, CI Therapy Research Group* (2011).

[25] David Webster and Ozkan Celik. 2014. Systematic review of Kinect applications in elderly care and stroke rehabilitation. *Journal of neuroengineering and rehabilitation* 11, 1 (2014), 108.