Classification of User Adherence to Home Hand Rehabilitation Technology Using a Feed-Forward Artificial Neural Network

Mohammad Shams
Department of Bioengineering
George Mason University
Fairfax, VA, USA
mshams4@gmu.edu

Daniel K. Zondervan
Flint Rehabilitation Devices LLC
Irvine, CA, USA
dzondervan@flintrehab.com

Quentin A. Sanders
Department of Bioengineering and Mechanical Engineering
George Mason University
Fairfax, VA, USA
qsanders@gmu.edu

***Abstract*—Hand impairments resulting from neurological conditions can significantly affect individuals' quality of life. Home-based rehabilitation programs are promising solutions to address these challenges. This study investigated user engagement with MusicGlove, a commercially available wearable grip sensor. We applied machine learning techniques to classify users based on their interaction with the device. We categorized users into 'low', 'moderate', and 'power' users and found considerable differences in device usage. For user adherence prediction after one day of device usage, we used a Multi-Layer Perceptron (MLP) deep learning model and traditional machine learning models such as K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Logistic Regression. The MLP model outperformed other models, achieving an average F1-score of 0.68 in cross validation and a balanced performance on unseen test data with an accuracy of 0.68, precision of 0.66, recall of 0.72, and an F1-score of 0.69 for the 'Low' user class. Our results underscore the need for personalized home-based rehabilitation programs and highlight the effective use of deep learning algorithms in predicting user adherence in home-based digital rehabilitation. This study contributes to the growing body of evidence supporting machine learning applications in healthcare, particularly in patient outcome prediction and treatment personalization.**

***Index Terms*—In-Home Rehabilitation, Wearable Sensor, Machine Learning, Patient Adherence**

I. Introduction

According to the World Health Organization (WHO), an estimated 255 million individuals live with a neurological injury such as stroke, spinal cord injury or traumatic brain injury [1]. Individuals living with these injuries often have impaired sensorimotor function which can significantly impact their overall quality of life [2]. To facilitate recovery of sensorimotor function, rehabilitation therapy is generally prescribed. However, the number of movement repetitions achieved during these therapy sessions is lower than what is thought necessary to facilitate the restoration of sensorimotor function [3]. Further, there have been several recent initiatives taken to shorten the duration of stay of individuals post neurological injury [4]. This potentially reduces cost, but also further limits the amount of therapy received in clinical settings.

Home-based therapy has been prescribed to increase the amount of therapy an individual achieves, although conventional home-based therapy has low adherence rates [5]. For example, Peiris et al. found that less than 50% of persons with stroke adhered to exercises prescribed in a home-based exercise program one month after hospital discharge [6]. Several factors possibly contribute to the low adherence rates observed such as lack of motivation, physical discomfort or perceived lack of time [7].

Wearable sensor technologies have been utilized to promote in-home rehabilitation amongst various patient populations [8]. Beyond providing objective and measurable data for monitoring adherence to rehabilitation exercises [9], [10], [11], these devices can be gamified to create an engaging therapeutic experience, potentially enhancing therapeutic effects. For example, our lab and others have used a wearable grip sensor to perform hand rehabilitation in the home setting with individuals in the chronic phase of spinal cord injury as well as individuals in the sub-acute, and chronic phase of stroke [12], [13]. In each of these different patient populations use of the wearable grip sensor resulted in improvements in hand function, and high levels of patient adherence compared to conventional home-based therapy options.

But, despite the promising outcomes demonstrated by prior research on wearable sensors, a comprehensive understanding of the device-related conditions that improve device usage is currently lacking. Power user analysis [14 can offer valuable insights in this regard. Power users demonstrate more innovative, efficient, and prolonged usage of a device's features compared to other users. By understanding the motivational characteristics of power users in the context of rehabilitation technology, it becomes possible to develop strategies that enhance user motivation, thereby increasing their movement practice at home.

There are several potential indicators to consider when studying power user behavior. Prior research has suggested that early success in device usage and parameter selection strategies can be key determinants of adherence. For instance, a study [15] investigated the link between challenge level and perseverance in unsupervised home rehabilitation using a sensor-based system, FitMe. Over two months individuals who achieved high, but not perfect success in the initial week of the exercise game demonstrated the greatest perseverance. Further, Sanders et al.'s study [9] of 10 subacute stroke patients using MusicGlove for home-based hand rehabilitation showed that users commonly adjust the game difficulty based on previous success, aiming to maintain high success rates in their practice.

In the present study, we analyzed usage logs from a large pool of anonymous users who utilized a commercially available device (MusicGlove) for at-home rehabilitation. Specifically, our objective with this large data set was to explore gaming-related features such as game success for example, that could differentiate power users from users with low device usage (low users). Afterwards, we utilized these features to develop and assess the accuracy of a machine learning algorithm designed to classify low users from power users.

II. Methods

A. MusicGlove Overview

MusicGlove, an FDA-listed medical device, is designed for aiding the recovery of individuals with hand impairments resulting from neurological injury [16]. The device is a wearable grip sensor, featuring six electrical leads strategically placed on the five fingertips and near the proximal interphalangeal joint on the lateral aspect of the index finger lateral (Figure 1).

MusicGlove allows users to perform five different types of grips — a key-pinch grip and opposition of the thumb to each of the four fingers. To operate the device individuals, touch the sensor on the thumb's tip to one of the other five sensors, in a manner that is coordinated with scrolling notes descending along a screen while music plays.

Additionally, the device offers three different song difficulty levels: easy, medium, or hard where increasing song difficulty increases the number of notes presented during the song, as well as the speed at which they descend along the screen.

Thus, users are able to adjust their training difficulty by changing the number of grip types needed to play each song, altering the song difficulty level, or adjusting both parameters. MusicGlove offers two different modes for users — "song" mode and "session" mode. In the "song" mode, users can modify game parameters after each song or replay the same song with the same parameters. The "session" mode, on the other hand, plays a sequence of songs with the song difficulty and number of grips used remaining unchanged, allowing users the option of ending the session early and modifying the parameters after each song.

Fig. 1. MusicGlove: An interactive device paired with a PC or tablet, featuring conductive fingertip pads to detect finger movements. It prompts users to perform timed grips in sync with scrolling notes descending on a computer screen.

The difficulty and number of grips used remain unchanged, allowing users the option of ending the session early and modifying the parameters after each song.

B. Data Cleaning

All users that were identified as test users, clinic users, or users with multiple software installations were removed. Additionally, entries from users with zero values for song duration (length of individual session measured in seconds), or notes presented (total number of notes shown during a song) were also excluded. We excluded users with zero 'cumulative grips' (the aggregate count of successful grip actions made by the user across sessions) to ensure the analysis involved only active MusicGlove users. Finally, we removed outliers, specifically entries falling below the 5% quantile for both song durations and notes presented, considering the resilience of quantiles against non-normal distributions. Following the application of these filters approximately 10% of users (173 in total) were eliminated, leaving 1,516 users.

C. Data Analysis

Device usage was visualized using a histogram. Further analysis of the histogram allowed for the categorization of the data into three distinct groups according to the level of activity: low users, moderate users, and power users. Low users were defined as those who used the device for less than 2 days, moderate users were individuals who used the device between 2-7 days, and power users were those who used the device for more than 7 days.

We generated descriptive statistics (mean and standard deviation) for the following metrics: day one success rate (number of notes completed / total number of notes presented on the first day of device use), difficulty level (# of grip types used + song difficulty with easy, medium, and hard corresponding to 1, 2, and 3 respectively), weighted score (difficulty level multiplied by success rate achieved by the user), the total number of grips hit, and the total number of sessions played. We performed independent two-tailed t-tests with the criteria for statistical significance set to $p < 0.05$ to compare differences between low users and power users. In our analysis, we were primarily interested in differentiating performance metrics between low users and power users. None of the statistical analyses performed included moderate users.

D. Machine Learning Model Selection and Development

We developed an array of ML models to classify users based on MusicGlove usage. All models were implemented in Python using the Google Colaboratory (Colab) platform. The developed models included a Multilayer Perceptron (MLP), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and a Logistic Regression model (LR). These models were selected because they have shown promising results in diverse set classification tasks [17].

The architecture of the MLP model consisted of two hidden layers with a Rectified Linear Unit (ReLU) activation function, which is a common choice due to its computational efficiency and ability to mitigate the vanishing gradient problem [20]. The model also utilized dropout regularization in between the hidden layers to prevent overfitting by randomly setting a fraction of input units to 0 during training [18]. The output layer used a sigmoid activation function, suitable for binary classification tasks.

Feature Selection: Each model developed in this study incorporated the following features: average first-day success rate, number of successful grips executed by each user, the total number of sessions, and the product of the number of games and the user's weighted score (success rate multiplied by game difficulty level).

The 'average success rate' provides a snapshot of initial user proficiency, the 'number of grip hits' reflects long-term engagement and skill improvement, the 'number of sessions played' indicates user engagement and commitment to therapy, and the 'sessions times weighted score' integrates user engagement and performance for a comprehensive assessment. The inclusion of these features was driven by their statistical significance (p-value < .0001 for each feature, low users vs. power users) in the current study, as well as their relevance in differentiating users in previous studies [15].

Model Training and Optimization: The models were trained using Adam, an adaptive learning rate optimization algorithm that is particularly effective for problems with large amounts of data or parameters [19]. The EarlyStopping callback from Keras was employed across all models to halt training when a monitored metric stopped improving, conserving computational resources, and minimizing overfitting risk.

Performance Evaluation: The performance of all developed models was evaluated using the F1 score and accuracy.

III. Results

A. MusicGlove Device Usage

After filtering, the sample consisted of 1,516 participants. These individuals performed an average of 84 ± 236 sessions spread over 11 ± 27 days. Each exercise had an average duration of 158.5 ± 28.7 seconds. This led to an average total exercise time of 221 ± 625 minutes per user. During this period, users completed an average of 7,936 ± 29,847 grip hits. Low users represented approximately 29.7% of the total amount of users while moderate, and power users represented 40.7% and 29.6% respectively.

Power users on average had a day one success rate of 65.36 ± 25.65 % compared to the 59.18 ± 29.06 % achieved by low users (p-value < .001, two-tailed t-test). Additionally, power users achieved a larger number of grips hits (3219.02 ± 7497.42 grips vs. 1253.45 ± 4779.31 grips, respectively, p-value < .0001, two-tailed t-test) and games played compared to low users (7.36 ± 6.51 sessions vs. 4.22 ± 3.80 sessions, respectively, p-value < .0001, two-tailed t-test).

B. Machine Learning Model Results

During model optimization, we explored various hyperparameters. We tested learning rates of 0.001, 0.005, 0.01, and 0.05; considered 32 and 64 neurons for the hidden layers; employed dropout rates of 0.5 and 0.6 to mitigate overfitting; and trained the model using batch sizes of 32 and 64 across both 100 and 200 epochs.

TABLE I

RESULTS FROM VARIOUS MACHINE LEARNING MODELS

Model	NN (MLP)	KNN	LR	SVM
Accuracy	**0.68**	0.63	0.67	0.59
Recall	**0.70**	0.57	0.67	0.57
Precision	0.66	**0.70**	0.67	0.65
F1-score	**0.68**	0.63	0.67	0.61

The optimal configuration of the MLP consisted of a learning rate of 0.01, a neural network of 32 neurons, a dropout rate of 0.5, a batch size of 32, and 200 training epochs. In comparative performance analysis, the MLP model achieved a greater level of classification accuracy in cross-validation trials when considering F1 score (MLP: 0.68, KNN: 0.63, LR: 0.67, SVM: 0.61, Table I). Further, it maintained balanced performance metrics on unseen test data, resulting in accuracy, precision, and recall of 0.66, 0.70, and 0.68 respectively for the 'Low' user class (Table I). The only category in which the MLP was outperformed was precision, with the KNN algorithm achieving a precision of 0.70 compared to the 0.68 achieved by the MLP.

IV. Discussions

A. Performance Metrics of Low Users versus Power Users

Several other rehabilitation studies have shown the importance of achieving high levels of success during the initial phase of therapy [20]. The results of this study are well aligned with the literature, as it was found that success on the first day of device usage was correlated with increased amounts of device usage. Power users also tended to play more games in comparison to low users. Several studies suggest that the amount of practice is maximized when the level of challenge is optimized. Thus, an interesting direction of future research would be to explore how power users were selecting parameters to modulate challenges, and subsequently increase their number of active days.

B. Classification of Device Usership Patterns Using Deep Learning

There have been several research groups that have attempted to utilize machine learning and deep learning approaches for rehabilitation approaches [21]. Here we developed a deep learning model as the first step towards developing an algorithm that can classify users with low device usage, and then subsequently alter game parameters to encourage continued use of the device.

We chose an artificial neural network for its adaptability and ability to detect nonlinear data relationships. While rule-based algorithms may seem straightforward, they lack the flexibility to capture evolving user behaviors or complex data interactions. Neural networks inherently understand these complexities and refine their insights with more data.

The optimized parameters of the deep learning model have resulted in satisfactory performance, as indicated by a balanced precision and recall reflected in the F1-score during cross validation, and a robust model evaluation result on unseen test data. This demonstrates the model's ability to identify and support low users, which aligns with the study's primary goal.

The precision and recall achieved by the model reveal a reasonably balanced trade-off, a factor of critical importance in this study. An equal emphasis on precision (how many of the users classified as 'low users' are truly low users) and recall (how many of the actual low users we managed to identify) ensures that the model correctly identifies 'low users' while minimizing the number of 'false positives'. This balance is particularly important in a healthcare context, where both over-prediction and under-prediction can have significant implications [22].

C. Limitations

The data in this study is anonymous, limiting our ability to analyze how demographic or clinical data influences device usage. Additionally, due to this anonymity, it is challenging to make any generalizations about the users, as they may not only be individuals with stroke but could also belong to different populations with impaired hand function. Further, while the MLP model boasts robust predictive capabilities, the interpretability of the model is challenging.

The modest F1 and overall accuracy scores reflect the challenges inherent in our research context. Our data, sourced from a diverse, unpredictable real-world environment, exhibits significant variability. These scores, while not ideal, provide a realistic baseline for gauging future model enhancements. As every user displays unique interaction styles and behavioral patterns, there may be additional or latent features not currently captured that could enhance classification performance. Future iterations of this work will focus on improving feature engineering and selection strategies to boost classification accuracy.

V. Conclusion

Utilizing the Multi-Layer Perceptron (MLP) machine learning model, we were successful in categorizing users based on their device engagement. This predictive capability can inform future intervention strategies aimed at enhancing user adherence, especially among 'low users,' by identifying individual-specific usage patterns and potential engagement barriers. The results from this study contribute to the growing body of evidence supporting the use of machine learning and deep learning methodologies in healthcare applications.

References

[1] A. Cieza, K. Causey, K. Kamenov, S. W. Hanson, S. Chatterji, and T. Vos, "Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: a systematic analysis for the Global Burden of Disease Study," *The Lancet*, vol. 396, no. 10267, 2006.

[2] S. L. Crichton, B. D. Bray, C. McKevitt, A. G. Rudd, and C. D. Wolfe, "Patient outcomes up to 15 years after stroke: Survival, disability, quality of life, cognition and mental health," *Journal of Neurology, Neurosurgery amp; Psychiatry*, vol. 87, no. 10, p. 1091–1098, 2016.

[3] C. E. Lang, "Observation of Amounts of Movement Practice Provided During Stroke Rehabilitation," *Archives of Physical Medicine and Rehabilitation*, vol. 90, no. 10, pp. 1692–1698, 2009.

[4] C. C. Chen, C. K. Chen, C. M. Tsai, C. H. Chang, L. W. Chou, and M. T. Lin, "Shortened hospital stay of patients with neurological injury in intensive care units," *Medicine*, no. 24, pp. 95–95, 2016.

[5] K. Jack, S. M. Mclean, J. K. Moffett, and E. Gardiner, "Barriers to treatment adherence in physiotherapy outpatient clinics: a systematic review," *Manual therapy*, vol. 15, no. 3, pp. 220–228, 2010.

[6] C. L. Peiris, N. F. Taylor, and N. Shields, "Extra physical therapy reduces patient length of stay and improves functional outcomes and quality of life in people with acute or subacute conditions: a systematic review," *Arch Phys Med Rehabil*, vol. 92, no. 9, pp. 1490–500, 2011.

[7] M. A. Holden, E. E. Nicholls, J. Young, E. M. Hay, and N. E. Foster, "Role of exercise for knee pain: what do older adults in the community think," *Arthritis care & research*, vol. 66, no. 10, pp. 1190–1198, 2014.

[8] P. Maceira-Elvira, T. Popa, A. C. Schmid, and F. C. Hummel, "Wearable technology in stroke rehabilitation: towards improved diagnosis and treatment of upper-limb motor impairment," *Journal of neuroengineering and rehabilitation*, vol. 16, no. 1, pp. 142–142, 2019.

[9] Q. Sanders, V. Chan, R. Augsburger, S. C. Cramer, D. J. Reinkensmeyer, and A. H. Do, "Feasibility of Wearable Sensing for In-Home Finger Rehabilitation Early After Stroke," *IEEE Trans. Neural Syst. Rehabil. Eng*, vol. 28, no. 6, pp. 1363–1372, 2020.

[10] F. Porciuncula, "Wearable Movement Sensors for Rehabilitation: A Focused Review of Technological and Clinical Advances," *PM&R*, vol. 10, no. 9S2, 2018.

[11] J. H. Burridge, "Telehealth, Wearable Sensors, and the Internet: Will They Improve Stroke Outcomes Through Increased Intensity of Therapy, Motivation, and Adherence to Rehabilitation Programs?" *J. Neurol. Phys. Ther*, vol. 41, pp. 32–38, 2017.

[12] Q. Sanders, V. Chan, R. Augsburger, S. C. Cramer, D. J. Reinkensmeyer, and K. Sharp, "Feasibility of home hand rehabilitation using musicglove after chronic spinal cord injury," *Spinal Cord Series and Cases*, vol. 8, no. 1, 2022.

[13] Q. Sanders, S. Okita, J. Lobo-Prat, D. S. De Lucena, B. W. Smith, and D. J. Reinkensmeyer, "Design and control of a novel grip amplifier to support pinch grip with a minimal soft hand exoskeleton," in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*, 2018, pp. 1089–1094.

[14] B. Zhong, "From smartphones to iPad: Power users' disposition toward mobile media devices," *Comput Human Behav*, vol. 29, no. 4, 2013.

[15] E. D. J. R. Muñoz, "Using Large-Scale Sensor Data to Test Factors Predictive of Perseverance in Home Movement Rehabilitation: Optimal Challenge and Steady Engagement," *Front. Neurol*, vol. 13, pp. 896298–896298, 2022.

[16] N. Friedman, V. Chan, D. Zondervan, M. Bachman, and D. J. Reinkensmeyer, "MusicGlove: motivating and quantifying hand movement rehabilitation by using functional grips to play music," *Annu Int Conf IEEE Eng Med Biol Soc*, vol. 2011, pp. 2359–63, 2011.

[17] S. Agatonovic-Kustrin and R. Beresford, "Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research," *J Pharm Biomed Anal*, vol. 22, no. 5, pp. 717–744, 2000.

[18] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[19] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," 2017.

[20] K. R. Lohse, C. G. Hilderman, K. L. Cheung, S. Tatla, and H. F. Van der Loos, "Virtual reality therapy for adults' post-stroke: A systematic review and meta-analysis exploring virtual environments and commercial games in therapy," *PLoS ONE*, vol. 9, no. 3, 2014.

[21] M. Jones, G. Collier, D. J. Reinkensmeyer, F. DeRuyter, J. Dzivak, D. Zondervan, and J. Morris, "Big data analytics and sensor-enhanced activity management to improve effectiveness and efficiency of outpatient medical rehabilitation," International Journal of Environmental Research and Public Health, vol. 17, no. 3, p. 748, 2020.

[22] X. Chen, H. Xie, F. L. Wang, Z. Liu, J. Xu, and T. Hao, "A bibliometric analysis of natural language processing in medical research," *BMC Med Inform Decis Mak*, vol. 18, no. 1, pp. 14–14, 2018.