A Low-Cost Wearable System to Support Upper Limb Rehabilitation in Resource-Constrained Settings

Md. Sabbir Ahmed[1(✉)], Shajnush Amir[1], Samuelson Atiba[2], Rahat Jahangir Rony[1], Nervo Verdezoto Dias[2], Valerie Sparkes[3], Katarzyna Stawarz[2], and Nova Ahmed[1]

[1] Design Inclusion and Access Lab (DIAL), North South University, Dhaka, Bangladesh
msg2sabbir@gmail.com, {shajnush.amir,rahat.rony, nova.ahmed}@northsouth.edu

[2] School of Computer Science and Informatics, Cardiff University, Cardiff, UK
{atibas,verdezotodiasn,stawarzk}@cardiff.ac.uk

[3] School of Healthcare Sciences, Cardiff University, Cardiff, UK
sparkesv@cardiff.ac.uk

Abstract. There is a lack of professional rehabilitation therapists and facilities in low-resource settings such as Bangladesh. In particular, the restrictively high costs of rehabilitative therapy have prompted a search for alternatives to traditional in-patient/out-patient hospital rehabilitation moving therapy outside healthcare settings. Considering the potential for home-based rehabilitation, we implemented a low-cost wearable system for 5 basic exercises namely, *hand raised, wrist flexion, wrist extension, wrist pronation, and wrist supination*, of upper limb (UL) rehabilitation through the incorporation of physiotherapists' perspectives. As a proof of concept, we collected data through our system from 10 Bangladeshi participants: 9 researchers and 1 undergoing physical therapy. Leveraging the system's sensed data, we developed a diverse set of machine learning models. And selected important features through three feature selection approaches: filter, wrapper, and embedded. We find that the Multilayer Perceptron classification model, which was developed by the embedded method Random Forest selected features, can identify the five exercises with a ROC-AUC score of 98.2% and sensitivity of 98%. Our system has the potential for providing real-time insights regarding the precision of the exercises which can facilitate home-based UL rehabilitation in resource-constrained settings.

Keywords: Upper limb rehabilitation · Low-resource · Wearable · Machine learning · Exercises · Physiotherapy · Bangladesh · Digital health · Low-cost wearable

Md. S. Ahmed and S. Amir—Equal contribution.

A. Tsanas and A. Triantafyllidis (Eds.): PH 2022, LNICST 488, pp. 33–45, 2023.
https://doi.org/10.1007/978-3-031-34586-9_3

1 Introduction

Upper limb impairment, a reduction or loss of limb function, is one of the most common consequences of acquired brain injury (ABI) [3]. In Bangladesh, ABI due to stroke and trauma is the leading cause of death and disability, representing an immense economic cost to the nation [20]. Over 97% of people with an ABI are diagnosed with some form of limb weakness that affects their ability to independently perform daily activities [3, 20]. In addition, there are very few care facilities and professionals [16, 20] limiting access to rehabilitation services, which increases the risk of long-term disability [20]. The high costs associated with hospital-based therapy continue to be a major barrier [20]. For example, in 2016, the typical Bangladeshi household income per month was 15,988 BDT ($189.76) [2] while the monthly cost for hospital-based rehabilitation in 2017 was 27,852 BDT ($328) [20]. High costs force one to choose between poverty and lifelong disability.

These challenges have created an opportunity for the use of technology to support home-based rehabilitation, especially in remote areas of Bangladesh. Technology-based rehabilitation in the home offers greater accessibility and convenience in relation to the time spent attending face-to-face appointments, thus reducing the overall costs of rehabilitation [17, 19]. Several technologies have been deployed for use in upper limb (UL) rehabilitation, including rehabilitation robots which actively assist patients to perform rehabilitation exercises [7], electrical stimulation which uses an electrical current to stimulate muscles in the affected limb [24], and wearable sensor devices which capture the patient's movements during rehabilitation exercises [17]. However, the robots are often large and expensive [19], and the electrical stimulation hardware requires expert knowledge and dexterous manipulation to set up [24]. Though there are several low-cost rehabilitation systems, there is a lack of computational models (e.g., [1, 11]) that could enable the systems to automatically identify the exercises and provide feedback to the patients and caregivers in real-time.

Therefore, we present a low-cost wearable system that incorporates machine learning models to support UL rehabilitation. Our contribution is twofold:

- We present a low-cost (around $16) system for recording and monitoring exercises to support UL rehabilitation.
- We develop machine learning (ML) models based on 14 algorithms and three feature selection (FS) approaches and show that the Multilayer Perceptron (MLP) model performed best with a ROC-AUC and precision score of 98.2%.

Overall, our system can facilitate home-based UL rehabilitation and real-time monitoring of the patients in low-resource settings.

2 Related Work

2.1 Approaches to Upper Limb Rehabilitation

Conventional rehabilitation is typically conducted in a controlled hospital environment. The methods for UL rehabilitation include mental imagery and action observation [12], constraint-induced movement therapy [14], and task-specific training [10]. Hospital-based task-specific training can lead to improved rehabilitation outcomes when administered frequently over an extended period [4, 10]. However, trained competencies acquired in the hospital environment, such as grasping and reaching, often fail to transfer to home and work environments, since trained movements may not correspond to activities in daily life [9].

Compared to hospital-based rehabilitation, home-based rehabilitation focused on everyday actions has been shown to achieve significantly better outcomes with regard to training transfer. This is because the training exercises are carried out within the relevant context where they would occur daily [21]. In addition, since home-based rehabilitation reduces the need for frequent hospital visits, and does not require expensive facilities, the cost of rehabilitation can be greatly lowered.

2.2 Technologies to Support Upper Limb Rehabilitation

In response to the demand for technological interventions for rehabilitation, several technologies have been deployed to support UL rehabilitation. These technologies include rehabilitation robots [7], electrical stimulation [24], and wearable sensor devices [17]. Rehabilitation robots, such as exoskeletons and soft wearable robots [7, 19], allow for precise movement control while providing assistance to weakened or paralyzed limbs during rehabilitation. But they are very expensive, not easily portable, and hard to wear and undress. Also, they often pose a safety risk when there is a misalignment between the robot and the human anatomy [19]. Consequently, rehabilitation robots are deployed in controlled hospital environments where the expertise is available to support clinical and rehabilitation practices. Thus, these technologies are often unsuitable for home-based rehabilitation.

Electrical stimulation (ES) for rehabilitation is focused on producing motor responses in muscles that are weakened or paralyzed due to an upper motor neuron injury, as is the case in people with ABI [24]. Its major setback in rehabilitation is the possibility of fatigue due to neurotransmitter depletion or propagation failure. When such fatigue sets in, the muscle fibers are not sufficiently stimulated and hence do not gain strength [24]. Therefore, expert knowledge is required to control the parameters of the stimulation provided. The need for expert monitoring and expensive specialized equipment restricts the deployment of ES in home-based rehabilitation settings.

Wearable devices are a widely explored system for UL rehabilitation [25]. They are lightweight, easy to put on and take off, cost-effective, and easy to operate [1, 17]. In addition, it is feasible to deploy them in the home as an alternative and/or complement to hospital-based rehabilitation [5]. Due to their low cost, they are also suitable for low-income settings. As such, researchers developed systems for the Global South focusing on exercises such as flexion, extension, abduction, horizontal abduction [1], supine [11], etc. However, they rely on visualization techniques which may not be precise enough to account for subtle differences to accurately identify UL exercises.

3 System Development

3.1 Understanding Physiotherapists' Perspectives

The developed system was informed by interviews with four Bangladeshi (3 men, 1 woman) physiotherapists, who helped us to identify basic exercises that were important for UL rehabilitation. To provide context, we summarize the key points that informed the design of our system; detailed interview results are reported elsewhere.

In Bangladesh, physiotherapy focuses on basic movements aiming to strengthen the muscles. However, the limited access to treatment was further worsened during the pandemic, as many centers closed down and physiotherapy sessions were discontinued:

"All patients' treatments do not complete at-home services. Sometimes there are required machines. So, that is not possible at home rather than in centers. In Bangladesh, good physio centers do not have many branches, so that people can't get support during COVID-19."- Physiotherapist 1.

This situation highlights the need for home-based physiotherapy. However, physiotherapists mentioned issues with the accuracy of movements when practicing at home, which can have a negative impact on patients:

"For hand movements, patients sometimes lift the wrong shoulder. Here movement is done but wrong. Detecting accurate movement is necessary" - Physiotherapist 4.

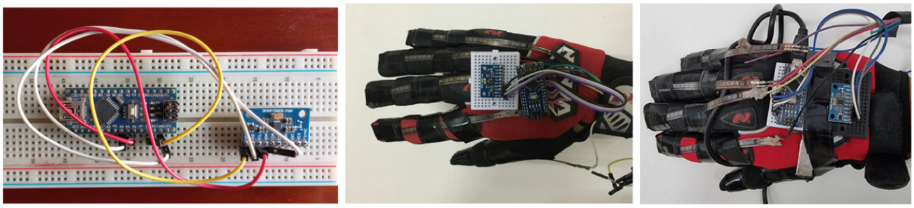
Maintaining accuracy at home requires a system that can monitor the patients' exercises. They have suggested key 5 exercises necessary for UL progress:

- *Hand raised*: It is an exercise where the hands are kept up 90° and the shoulders are kept straight.
- *Wrist flexion*: It is the bending of the hand down at the wrist where the palm faces toward the arm.
- *Wrist extension*: It is the opposite of flexion where the movement of the hand is backward, towards the forearm's posterior side.
- *Wrist pronation*: In pronation, the forearm or palm faces down.
- *Wrist supination*: In this exercise, the forearm or palm faces up.

3.2 System Design

To develop a low-cost system that can facilitate the identification of the aforementioned exercises unobtrusively, we used Arduino Nano (price ~$7) and inertial measurement unit (IMU) sensor MPU-9250 9-DOF (price ~$9) where the IMU consists of an accelerometer, gyroscope, and magnetometer. Firstly, a basic prototype (Fig. 1(a)) was developed to ensure the component level accuracy, which was followed by a working prototype (Fig. 1(b)) on a glove that had a flex sensor placed on each finger. However, the sensors' placement added extra noise which was finally modified (Fig. 1(c)) by keeping the sensors further from the finger.

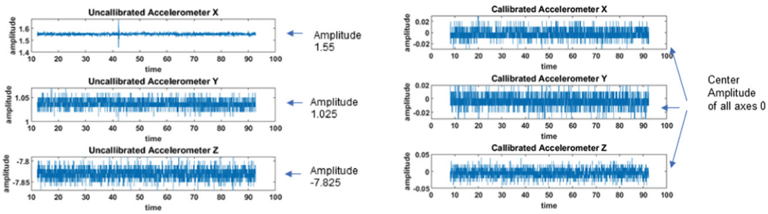
Data Recording Application: A data recorder was developed to prompt the user to record the data of each exercise. The user could control the beginning of the recording



(a) Basic prototype (b) Working prototype (c) Final design

Fig. 1. Development of the system.

of each exercise. In the data recorder, the delay between each movement of the same exercise can be modified as the users may have different preferences.



(a) Uncalibrated data (b) Calibrated data

Fig. 2. (a) Uncalibrated and (b) calibrated data of the accelerometer.

Initialization and Stabilization: The sampling rate of the system was 90 Hz. The raw data of the accelerometer, as an example, portrayed in Fig. 2(a), is taken before any hand gesture takes place to get a sense of the IMU sensor's data. As seen in Fig. 2(a), the starting amplitude without any hand gestures is non-zero. The uncalibrated values are due to gravity, the earth's magnetic flux coupled with the sensor's offset values which implies that the sensors' starting value will be different in different starting positions of the sensor. To ensure amplitude values are stabilized, a calibration routine was implemented where during the initialization of the wearable at a flat surface, 3000 readings are taken and averaged. This averaged value is subtracted from all subsequent sensor data to remove the offset. The data after the calibration is illustrated in Fig. 2(b). Due to the calibration routine, the amplitude of all 3 axes of data points is always set to zero at the initiation of the device for rest position.

Precision Adjustment: Initially, there was low precision as the data recorder registered integer values to reduce processing time. Due to the absence of decimal points, the data plots were not continuous (Fig. 3). To get the decimal points, we multiplied the sensor data by 100,000 and divided the integer data by that value, thereby incorporating the lost decimal values and increasing the precision. E.g., for the low precision gyroscope data, there is a staircase effect, but for the high precision data, the plot line is continuous as decimal points are incorporated (Fig. 3).

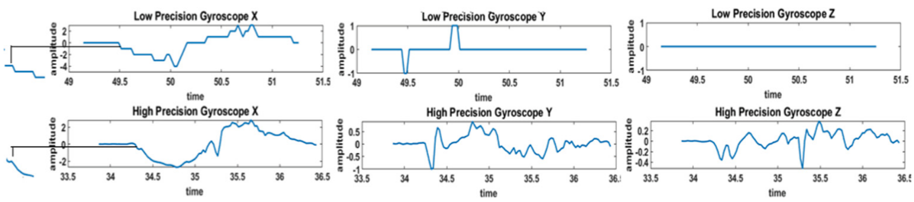


Fig. 3. Improved data precision of the gyroscope.

3.3 Validation of the Developed System

For validation, our system's retrieved data was compared with a Samsung Galaxy S6 smartphone (Table 1) which has been found as a promising device to sense data [18]. To get the sensors' data from Galaxy S6, we used the Physics *Toolbox Sensor Suite Pro* app which is available in the Play Store. We recruited 2 participants and each participant performed each exercise through our system and also through Galaxy S6. During validation, both the prototype and the smartphone were connected to the user at the same time and the data was captured in both devices simultaneously. Therefore, both sets of data captured the same exercise movement data.

Table 1. Comparison of our system's retrieved data with the Galaxy S6.

Axis	Accelerometer			Gyroscope			Magnetometer		
	Peak diff	Noise variance		Peak diff	Noise variance		Peak diff.	Noise variance	
		Our system	Galaxy S6		Our system	Galaxy S6		Our system	Galaxy S6
X	8.98%	$5.7 * 10^{-7}$	0.0942	1.14%	$4.58 * 10^{-5}$	$9.42 * 10^{-5}$	1.14%	$4.58 * 10^{-5}$	$9.42 * 10^{-5}$
Y	5.98%	$3.76 * 10^{-7}$	$2.07 * 10^{-4}$	0.58%	$6.11 * 10^{-8}$	$1.05 * 10^{-4}$	0.58%	$6.12 * 10^{-8}$	$1.05 * 10^{-4}$
Z	1.22%	$8.12 * 10^{-5}$	$8.19 * 10^{-4}$	16.17%	$2.47 * 10^{-7}$	$1.24 * 10^{-5}$	16.17%	$2.47 * 10^{-7}$	$1.24 * 10^{-5}$

For comparison, we calculated the noise variance using the MATLAB function *evar* [8] which estimates the variance of additive noise. Peak difference was calculated using the formula $peak\_dif = \frac{|(peak\_value_{our\_syetm} - peak\_value_{S6})|}{peak\_value_{S6}}$. In the Z-axis of the accelerometer and the X-axis of the gyroscope and magnetometer, the peak difference between our system and S6 retrieved data was less than 1.5% (Table 1). Though in some cases the peak difference was around 16%, our system had much lower noise variance. Thus, our system's data was comparable to the S6 phone with better noise performance.

4 Methodology

4.1 Participants and Research Ethics

As a proof of concept, we conducted a study in Bangladesh with 10 participants, nine researchers, and one undergoing physiotherapy. On average, each participant provided data on 10.8 different days (SD = 7.20, Minimum = 3, Maximum = 30, Median = 10), and on each day, each participant did each exercise 5 times. While collecting data, we labeled the 5 exercises (e.g., wrist flexion) so that the ML models' prediction could be evaluated.

The study was approved by the North South University IRB/ERC committee (2020/OR-NSU/IRB-No.0501). We received the participants' signed consent forms.

4.2 ML Model Development

4.2.1 Feature Extraction and Selection

For each participant, we calculated 6 types of data, namely, mean, standard deviation (SD), interquartile range (IQR), skewness, kurtosis, and entropy over the time periods based on the accelerometer, magnetometer, and gyroscope sensed data from each of the 3 axes (X, Y, Z) separately. In total, we extracted 54 features (6 types of data * 3 sensors * 3 axes) from each participant. But 36 (67%) of the features' data were not normally distributed, and thus, we normalized the data instead of standardization.

In general, feature selection (FS) methods can be grouped into 3 categories [15]: wrapper, filter, and embedded method. As a wrapper method, we used the Boruta algorithm which is an all-relevant FS approach [23]. We tuned the maximum depth of Boruta's base estimator Random Forest (RF) algorithm and the range was 3 to 7 which is suggested to use [6]. We used the Information Gain (IG) and RF algorithms as the filter and embedded methods respectively. IG and RF algorithms work by a minimal-optimal method whereas, unlike the all-relevant FS approach, it does not inform a fixed set of features to be used. Therefore, for the IG and RF methods, we used the maximum length of the Boruta selected features set as the upper boundary and 1 as the lower boundary of the number of features to be selected.

4.2.2 Model Development and Validation

Based on the "No Free Lunch" theorem, there is no algorithm that can perform best for all problems. Hence, we developed models (Fig. 4) by exploring a diverse set of ML algorithms: Logistic Regression (Logit), K-Nearest Neighbor (KNN), Support Vector Classifier (SVC), Gaussian Naïve Bayes (GNB), Decision Trees, Random Forest (RF), Gradient Boosting (GB), Light GBM, AdaBoost, Extra Tree, CatBoost, Extreme Gradient Boosting, and Multilayer Perceptron (MLP). In addition, a Dummy classifier was used as the baseline which predicts regardless of the input features. Inspired by Vabalas et al. [22], we used the nested cross-validation approach which shows generalizable performance. In the outer loop, there was Leave One Out Cross Validation (LOOCV) where we divided the dataset into n equal portions where each portion presenting a participant's data. Then, we used $n - 1$ participants' data to select the best set of features, and in the inner loop, to tune the hyper-parameters, we used a 5-fold CV maximizing the macro-F1 score. For tuning, we used the Bayesian search technique. After finding the best estimator, we predicted the class of the left participant's exercises who was not involved in FS and hyper-parameter tuning stages. We repeated this process to predict the exercise class of each of the 10 participants.

We evaluated the models' performance by comparing the labeled class with the models' predicted exercise and calculated the precision, sensitivity, F1, and ROC-AUC (Area Under the Receiver Operating Characteristic Curve). Each evaluation metric's

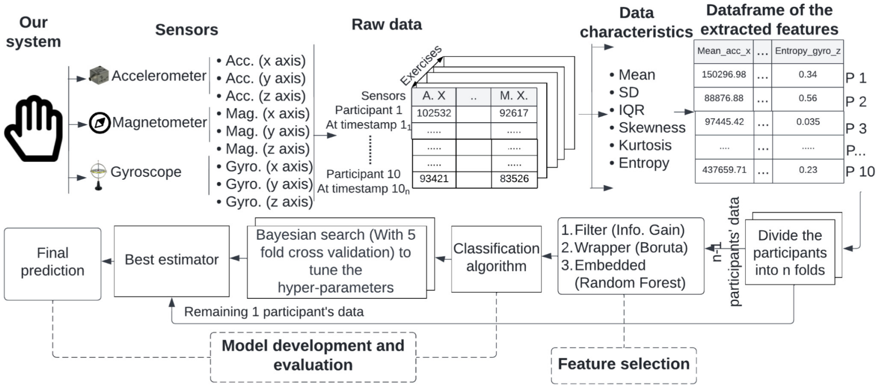
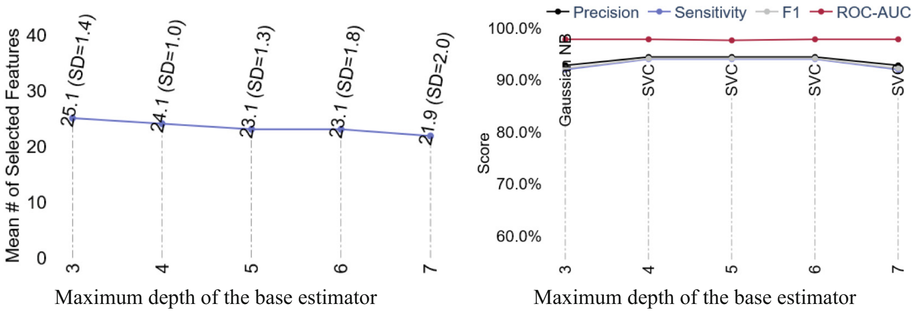


Fig. 4. ML pipeline to identify each exercise.

score was macro-averaged by calculating the simple arithmetic mean of all the 5 class scores of the evaluation metric (e.g., precision). It should be noted that in our study, each participant performed each of the 5 exercises, which means there is no class imbalance. In addition, to make the models unbiased, none of the 5 exercise data of the test participants were used in the training phase.

5 Results

5.1 Predicting the Exercises



(a) Number of selected features

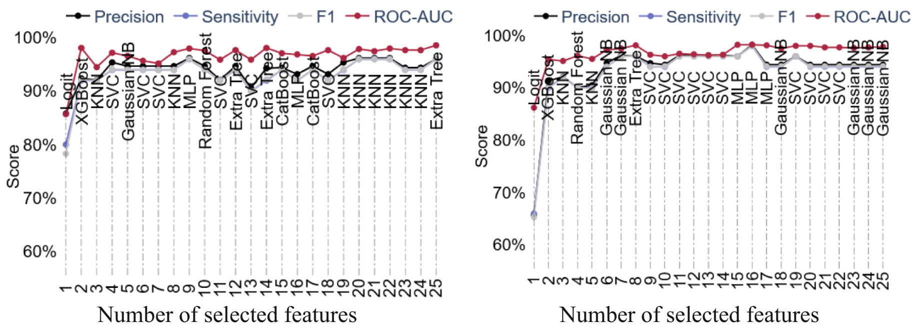
(b) Best models' performance

Fig. 5. (a) Number of selected features and (b) performance of the best model while selecting the features by tuning the maximum depth of the base estimator of Boruta.

To identify each exercise from the sensor retrieved data, we tuned the maximum depth of the base estimator Random Forest (RF) in the all-relevant FS approach Boruta. The number of selected features was lower with the increase in the maximum depth (Fig. 5(a)). It is apparent that at maximum depths 4, 5, and 6, the best performing model SVC has almost identical performance (Precision = 94.4%, F1 = 94.1%, Sensitivity

= 94%) where SVC can identify 94% of exercises accurately (Fig. 5(b)). However, the mean number of selected features in each iteration of LOOCV is 24.1 (SD = 1.0) at depth 4 whereas it is 23.1 at depths 5 and 6 (Fig. 5(a)). As the ROC-AUC score is 97.6% and 97.8% at depths 5 and 6 respectively, we consider SVC at depth 6 as the best model due to having relatively higher predictability.

In the Boruta FS approach, on average, the maximum number of selected features was 25.1 which is at depth 3 (Fig. 5(a)). Therefore, as discussed in Sect. 4.2.1, we set 25 as the upper boundary of the number of features to be selected in the filter and embedded FS methods. In the filter method Information Gain (IG), when there is only 1 feature selected, the Logit model performed best with an F1 score of around 80% (Fig. 6(a)). However, the MLP model based on 9 important features in each iteration of LOOCV had a maximum F1 of 96% and a ROC-AUC score of 98%. Though at features 9, 20, 21, 22, and 25 the performance of the best model is almost similar, the model based on 9 features were selected as best due to having lower features.



(a) IG (b) RF

Fig. 6. Best models' performance when a number of important features are selected through the (a) filter method IG and (b) embedded method RF algorithm.

In the embedded method RF selected 9 important features, the best model SVC had a precision of 94.7%, an F1 score of 94%, and a ROC-AUC score of 96% (Fig. 6(b)) which was lower than the performance of the best model based on IG selected 9 features. However, the MLP model based on the RF selected 16 features in each iteration of LOOCV has a ROC-AUC score of 98.2%, precision of 98.2%, and F1 score of 98%. This MLP model had higher performance than any other models based on the Boruta (Fig. 5) and IG selected features (Fig. 6(a)).

Though we developed models based on 14 algorithms, we found conventional ML models' (e.g., SVC, GN) superior performance. In the Boruta selected features, the best model in each depth was either SVC or the GNB (Fig. 5(b)). Also, in IG (Fig. 6 (a)) and RF (Fig. 6(b)) selected feature-based models, the best performing model in most cases was KNN and SVC. Apart from these, though there was a single tree-based model among the top-5 models in the case of each FS method's best set of features, we found 3 conventional algorithm-based models (Table 2) which shows their robustness to identify

Table 2. Performance of the top-5 classifiers and baseline Dummy classifier, based on the best (in terms of ML models' performance) set of features of each FS method. "# of features" present the number of features used in each iteration of LOPOCV. E: Extra.

Filter method IG (# of features = 9)					Wrapper method Boruta (average # of features = 23.1 (SD: 1.8))					Embedded method RF (# of features = 16)				
Model Name	Precision	Sensitivity	F1	ROC AUC	Model Name	Precision	Sensitivity	F1	ROC AUC	Model Name	Precision	Sensitivity	F1	ROC AUC
MLP	96.2	96	96	98	SVC	94.4	94	94.1	97.8	MLP	98.2	98	98	98.2
SVC	93.2	92	92	96.8	GNB	92.8	92	92.2	97.8	GNB	94.4	94	94.1	97.8
Logit	90	90	90	96	MLP	91	90	90.1	95.9	SVC	94	94	93.9	98
E. Tree	90	90	90	95.9	E. Tree	90.4	90	90.1	98.3	KNN	92.3	90	89.7	93.8
GNB	90.1	90	90	96.8	Logit	89.9	90	89.9	97.1	RF	88.6	88	88	95.6
Dummy	0	0	0	50	Dummy	0	0	0	0	Dummy	0	0	0	50

the exercises. We also found in each FS method, the models had higher scores compared to the baseline dummy classifier (Table 2).

While exploring more the performance of the best classifier regardless of FS method, we found that the MLP model identified hand raised, wrist pronation, and supination 100% accurately (Table 3). However, in wrist flexion, though the predicted class was 100% accurate, it correctly identified 90% of exercises among 10 flexion exercises of 10 participants (precision = 100%, sensitivity = 90%, support = 10).

Table 3. Best model's (MLP based on 16 features selected by RF) prediction for each exercise.

Exercise	Precision	Sensitivity	F1	Support	Exercise	Precision	Sensitivity	F1	Support
Hand raised	100	100	100	10	Wrist pronation	100	100	100	10
Wrist flexion	100	90	95	10	Wrist supination	100	100	100	10
Wrist extension	91	100	95	10					

5.2 Feature Importance

We found 29 features (Fig. 7) that were used in the top-5 classifiers on the basis of the best set of features of each FS method (Table 2). Among them, 14 features (48.28%) were based on the gyroscope sensed data, which reflects that this sensor's features are more important for identifying the exercises (Fig. 7).

In the Boruta and RF FS methods, we found the stability of 6 features such as the mean and skewness of the gyroscope sensed data in the Z-axis, which appeared in all iterations of the LOOCV (Fig. 7). This may explain the fact of having relatively identical performance in RF and Boruta selected feature-based ML models. For example, the best model based on the RF selected features from 11 to 14 and also from 20 to 25, had identical performance (Fig. 6(b)). Also, at depths 4, 5, and 6 of the base estimator of Boruta FS, the performance was almost identical (Fig. 5(b)).

Feature Name	Info. Gain	Boruta	Random Forest	All 3 FS Methods
SD A_Z	100	100	100	100
Mean A_Z	100	100	90	96.7
Mean G_Z	70	100	100	90
IQR M_Z	60	100	100	86.7
SD M_Z	90	100	60	83.3
IQR G_Y	30	100	100	76.7
Skewness G_Z	20	100	100	73.3
Mean A_Y	20	100	100	73.3
SD G_X	0	100	100	66.7
Skewness G_X	0	100	50	50
Entropy G_Y	20	100	30	50

Feature Name	Info. Gain	Boruta	Random Forest	All 3 FS Methods
Kurtosis G_X	0	100	30	43.3
Mean M_Y	0	100	20	40
Mean M_Z	0	80	40	40
IQR M_X	10	90	20	40
Skewness G_Y	0	80	30	36.7
SD G_Y	0	100	0	33.3
Kurtosis A_Y	0	40	50	30
IQR G_X	40	20	10	23.3
Skewness A_X	0	20	20	13.3
Kurtosis G_Z	0	10	10	6.7
IQR A_X	0	10	0	3.3
SD M_X	0	10	0	3.3

Fig. 7. Features used to develop the top-5 classifiers. Here, each value presents each feature's percentage of appearance in all 10 iterations of LOOCV. A: Accelerometer, M: Magnetometer, G: Gyroscope. X, Y, and Z denote the axes.

6 Discussion

We presented a low-cost system (~$16) to support UL rehabilitation in resource-constrained settings. Based on our system's sensed data, we developed ML models which can identify the 5 exercises with a ROC-AUC score of over 95%. This extends the existing systems, particularly the low-cost systems focusing on a few other exercises [1] where there is no automated process for accurate identification of exercises [1, 11]. Therefore, our system could identify and inform patients and caregivers whether the particular exercise is conducted precisely. This could facilitate home-based rehabilitation and support physiotherapists in remote monitoring, especially when there are inadequate rehabilitation facilities [16].

We found MLP as the best-performing model where its predicted exercise was accurate in 98.2% of cases. A plausible reason for the higher performance of MLP can be due to the neural networks' ability to capture complex patterns. But recent systematic reviews in medical informatics found researchers' preference for tree-based ML algorithms [13]. Though we developed models based on 8 tree-based algorithms, in the top-5 classifiers of each FS method, we found a single tree-based model. However, there were 3 models based on conventional ML algorithms such as the SVC and Logit where evaluation metrics' scores were over 90%. Conventional ML algorithms have fewer parameters that do not get overfitted easily. Also, considering the smaller sample size, we used nested cross-validation to build the models, which are found to prevent overfitting and show unbiased performance [22]. Hence, our findings suggest incorporating conventional ML algorithms along with complex algorithms while developing models to identify exercises for UL rehabilitation.

7 Limitations

The main limitations of our study are the low number of participants, especially with impairments in the upper extremities or undergoing physical rehabilitation. As the aim of this study was to evaluate the feasibility of our proof of concept system, future work should focus on applying more robust evaluation methods.

8 Conclusion

We presented an affordable system that was designed by integrating physiotherapists' perspectives. We presented the applicability of our system in accurately identifying 5 exercises with 10 participants to show its feasibility. Our system can play a role in home-based UL rehabilitation in low-resource settings such as Bangladesh.

Acknowledgment. We thank Dr. Catt Turney for input in the physiotherapist research, and all participants for their time. This study was supported by seed funding from Cardiff University's GCRF QR Funding from the Higher Education Funding Council for Wales.

References

1. Anowar, J., Ali, A.A., Amin, M.A.: A low-cost wearable rehabilitation device. In: Proceedings of the 2020 12th ICCAE. ACM (2020)
2. Bangladesh Bureau of Statistics. Household Income and Expenditure Survey 2016
3. Chakraborty, P.K., Islam, M.J., Hossain, M.S., Barua, S.K., Rahman, S.: Profile of patients receiving stroke rehabilitation in A tertiary care Hospital. Chattagram Maa-O-Shishu Hosp. Med. Coll. j. **17**, 9–12 (2018). https://doi.org/10.3329/cmoshmcj.v17i1.39435
4. D'Auria, D., Persia, F., Siciliano, B.: Human-computer interaction in healthcare: how to support patients during their wrist rehabilitation. In: 2016 IEEE Tenth ICSC. IEEE (2016)
5. Dutta, D., Sen, S., Aruchamy, S., Mandal, S.: Prevalence of post-stroke upper extremity paresis in developing countries and significance of m-Health for rehabilitation after stroke - a review. Smart Health **23**, 100264 (2022). https://doi.org/10.1016/j.smhl.2022.100264
6. boruta_py. https://github.com/scikit-learn-contrib/boruta_py. Accessed 03 Aug 2022
7. Tran, P., Jeong, S., Wolf, S.L., Desai, J.P.: Patient-specific, voice-controlled, robotic FLEXotendon glove-II system for spinal cord injury. IEEE Robot. Autom. Lett. (2020)
8. EVAR - Noise variance estimation. https://www.biomecardio.com/matlab/evar_doc.html. Accessed 05 Aug 2022
9. Grossman, R., Salas, E.: The transfer of training: what really matters: the transfer of training. Int. J. Train. Dev. **15**, 103–120 (2011)
10. Hubbard, I.J., Parsons, M.W., Neilson, C., Carey, L.M.: Task-specific training: evidence for and translation to clinical practice: task-specific training in clinical practice. Occup. Ther. Int. **16**, 175–189 (2009). https://doi.org/10.1002/oti.275
11. Hughes, C.M.L., et al.: Development of a post-stroke upper limb rehabilitation wearable sensor for use in sub-Saharan Africa: a pilot validation study. Front. Bioeng. Biotechnol. **7**, 322 (2019)
12. Ietswaart, M., et al.: Mental practice with motor imagery in stroke recovery: randomized controlled trial of efficacy. Brain **134**, 1373–1386 (2011)

13. Brnabic, A., Hess, L.M.: Systematic literature review of machine learning methods used in the analysis of real-world data for patient-provider decision making. BMC Med. Inform. Decis. Mak. **21**, 54 (2021). https://doi.org/10.1186/s12911-021-01403-2
14. Sunderland, A., Tuke, A.: Neuroplasticity, learning and recovery after stroke: a critical evaluation of constraint-induced therapy. Neuropsychol. Rehabil. **15**, 81–96 (2005)
15. Guyon, I., Nikravesh, M., Gunn, S., Zadeh, L.A. (eds.): Feature Extraction: Foundations and Applications. Springer, Heidelberg (2006)
16. Uddin, T., Islam, M.T., Rathore, F.A., O'Connell, C.: Disability and rehabilitation medicine in Bangladesh: current scenario and future perspectives. J. Int. Soc. Phys. Rehabil. Med. (2019)
17. Low, K.S., Lee, G.X., Taher, T.: A wearable wireless sensor network for human limbs monitoring. In: 2009 IEEE I2MTC. IEEE (2009)
18. Hsieh, K.L., Sosnoff, J.J.: Smartphone accelerometry to assess postural control in individuals with multiple sclerosis. Gait Posture **84**, 114–119 (2021)
19. Maciejasz, P., Eschweiler, J., Gerlach-Hahn, K., Jansen-Troy, A., Leonhardt, S.: A survey on robotic devices for upper limb rehabilitation. J. Neuroeng. Rehabil. **11**, 3 (2014)
20. Mamin, F.A., Islam, M.S., Rumana, F.S., Faruqui, F.: Profile of stroke patients treated at a rehabilitation centre in Bangladesh. BMC Res. Notes **10**, 520 (2017)
21. Mawson, S., Nasr, N., Parker, J., Davies, R., Zheng, H., Mountain, G.: A personalized self-management rehabilitation system with an intelligent shoe for stroke survivors: a realist evaluation. JMIR Rehabil. Assist. Technol. **3**, e1 (2016)
22. Vabalas, A., Gowen, E., Poliakoff, E., Casson, A.J.: Machine learning algorithm validation with a limited sample size. PLoS ONE **14**, e0224365 (2019)
23. Kursa, M.B., Rudnicki, W.R.: Feature selection with the Boruta package. J. Stat. Softw. (2010). https://www.jstatsoft.org/article/view/v036i11
24. Electrical Stimulation - Its role in upper limb recovery post-stroke. https://www.physio-pedia.com/index.php?title=Electrical_Stimulation_-_Its_role_in_upper_limb_recovery_post-stroke&oldid=216559. Accessed 01 Aug 2022
25. Maceira-Elvira, P., Popa, T., Schmid, A.-C., Hummel, F.C.: Wearable technology in stroke rehabilitation: towards improved diagnosis and treatment of upper-limb motor impairment. J. Neuroeng. Rehabil. **16**, 142 (2019). https://doi.org/10.1186/s12984-019-0612-y