Quantification of Motor Function Post-Stroke Using Novel Combination of Wearable Inertial and Mechanomyographic Sensors

Lewis Formstone, Weiguang Huo, Samuel Wilson, Alison McGregor, Paul Bentley, and Ravi Vaidyanathan

***Abstract*— Subjective clinical rating scales represent the gold-standard for diagnosis of motor function following stroke. In practice however, they suffer from well-recognized limitations including assessor variance, low inter-rater reliability and low resolution. Automated systems have been proposed for empirical quantification but have not significantly impacted clinical practice. We address translational challenges in this arena through: (1) implementation of a novel sensor suite combining inertial measurement and mechanomyography (MMG) to quantify hand and wrist motor function; and (2) introduction of a new range of signal features extracted from the suite to supplement predicted clinical scores. The wearable sensors, signal features, and machine learning algorithms have been combined to produce classified ratings from the Fugl-Meyer clinical assessment rating scale. Furthermore, we have designed the system to augment clinical rating with several sensor-derived supplementary features encompassing critical aspects of motor dysfunction (e.g. joint angle, muscle activity, etc.). Performance is validated through a large-scale study on a post-stroke cohort of 64 patients. Fugl-Meyer Assessment tasks were classified with 75% accuracy for gross motor tasks and 62% for hand/wrist motor tasks. Of greater import, supplementary features demonstrated concurrent validity with Fugl-Meyer ratings, evidencing their utility as new measures of motor function suited to automated assessment. Finally, the supplementary features also provide continuous measures of sub-components of motor function, offering the potential to complement low accuracy but well-validated clinical rating scales when high-quality motor outcome measures are required. We believe this work provides a basis for widespread clinical adoption of inertial-MMG sensor use for post-stroke clinical motor assessment.**

***Index Terms*— Stroke, Fugl-Meyer assessment, automated upper-limb assessment, wearables, machine learning, mechanomyography.**

I. Introduction

STROKE is the second largest cause of death globally and the second biggest cause of years lost prematurely or living with disability [1]. The standard clinical rating scales of motor function for stroke are widely used for monitoring subject improvement and defining rehabilitation requirements. These rating scales also form the gold standard method of quantifying motor function in research applications. Common applications include the assessment of new rehabilitation programmes [2], medications [3], and lesion-symptom mapping studies [4].

One of the most commonly used and widely validated [5] standard clinical rating scales is the Fugl-Meyer Assessment (FMA) [6]. A subsection of this scale is dedicated to the assessment of the upper extremity only (FMA-UE). This section examines each component of the upper extremity in isolation as well as combined through synergistic and non-synergistic movements. Each motor component/task is assigned a qualitative rating depending on how well it was performed with a range from 0 to 2. This range covers no movement/function (0), partial movement/function (1), and full movement/function (2).

Despite the global reliance on the standard clinical rating scales for rehabilitation and research, these methods suffer from several limitations such as the subjectivity of assessment, time required, and low resolution. To address one or more of these limitations, sensor-based "automated" systems have been proposed as an alternative method of quantifying motor function post-stroke. The majority of proposed sensors in prior automated systems have measured the kinematics of movements as measured either via camera-based [7]–[9] or wearable inertial-based [10]–[12] systems.

Manuscript received November 23, 2020; revised May 6, 2021; accepted June 6, 2021. Date of publication June 15, 2021; date of current version June 24, 2021. This work was supported in part by the U.K. Research Institute Engineering and Physical Sciences Research Council (UKRI EPSRC) Centre for Doctoral Training in Neurotechnology for Life and Health, in part by the Dementia Research Institute Care Research Technology Centre (DRI-CRT) under Grant UKDRI-7003, in part by the U.K. EPSRC under Grant EP/K503733/1, in part by the U.K. National Institute of Health Research Imperial Biomedical Research Centre (NIHR BRC), and in part by Serg Technologies, U.K. *(Corresponding author: Weiguang Huo.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Health Research Authority of the National Health Service (NHS), U.K., under Approval No. 11/LO/0941.

Lewis Formstone and Weiguang Huo are with the Department of Mechanical Engineering and DRI-CRT, Imperial College London, London SW7 2AZ, U.K. (e-mail: l.formstone15@imperial.ac.uk; w.huo@imperial.ac.uk).

Samuel Wilson and Ravi Vaidyanathan are with the Department of Mechanical Engineering and DRI-CRT, Imperial College London, London SW7 2AZ, U.K., and also with Serg Technologies, London E16 2DQ, U.K. (e-mail: s.wilson14@imperial.ac.uk; r.vaidyanathan@imperial.ac.uk).

Alison McGregor is with the Department of Surgery and Cancer, Imperial College London, London W12 0BZ, U.K. (e-mail: a.mcgregor@imperial.ac.uk).

Paul Bentley is with the Department of Brain Sciences, Imperial College London, Charing Cross Hospital, London W6 8RF, U.K. (e-mail: p.bentley@imperial.ac.uk).

Digital Object Identifier 10.1109/TNSRE.2021.3089613

A major limitation of prior post-stroke automated systems is that they fail to effectively quantify the movements of the hand or wrist, a major component of stroke dysfunction. The camera-based systems utilised in earlier trials have shown poor performance when tracking these regions [13]. Inertial-based sensors may perform better but require many sensors and careful placement to capture each finger joint. One solution to quantifying the hand-region has been proposed in the form of instrumented gloves [14], [15] but these have limitations including poor fit and hygiene concerns.

An emerging modality of interest for measuring hand and wrist motor function is myographic data. This modality has shown good potential for distinguishing stroke impairment in the literature [16], [17]. In addition, myographic sensing may be used to capture the activity of muscles in the lower arm which have a functional role in the hand and wrist. The application of electromyography (EMG) has already been shown to perform well when classifying hand and wrist function as part of an automated sensor system [18], [19]. However, the application the EMG in such sensor systems is limited in practice by factors including requirements for careful skin preparation and electrode placement, and a signal which is dependent on humidity and skin impedance.

The system developed in the present study proposes the novel inclusion of mechanomyography (MMG) as an alternative to EMG. MMG is less widely validated than EMG for clinical applications but has several advantages that lend itself for use in a wearable system. These include re-usability, higher signal-to-noise ratio, possessing a robust signal which is more independent to changes of skin impedance, and ease of application [20]. Our past work has implemented MMG to diagnosis motor dysfunction in Parkinson's disease [21], quantify hand function [22], fabricate clothing for stroke telemedicine [23] and track of arm movement and muscle activity in parallel [24]. MMG will be positioned in the present study to record hand and wrist function by calculating features of the signal which have shown to be statistically different based on stroke severity [25].

There are three main methods of calculating motor function that have been used in existing automated systems. The first method is by standard classification or regression models which are trained using labels provided by clinician-assessed clinical rating scales [8], [11]. Secondly, custom logic-based classification models have been developed using domain knowledge to design the model [7], [13]. Finally, algorithmic methods have been devised for outputting novel metrics of motor function [26], [27].

Classification or regression models of motor clinical rating scales have the advantage of classifying to a clinically relevant and well-validated score. Disadvantages include the label noise introduced by limited rater reliability, the low resolution of labels, and the requirement for a large volume of clinical data. The application of logic-based classification mitigates the requirements for a large volume of clinical data but shares the same low-resolution of the clinical rating scale predicted.

Algorithmic methods of evaluating motor function derive novel scores without requiring an external rating score. These methods have been proposed as a means of calculating a score with a resolution which far exceeds that of the standard clinical rating scales. The primary limitation of prior devised algorithmic methods is that they compute a score by comparison of the "affected" and "unaffected" upper limbs post-stroke. This method is dependent on the subject suffering from motor dysfunction to only one-side (hemiplegia) which is not always present post-stroke. In addition, assessment of both arms of the subject is required which doubles the testing duration.

The study by Song *et al* [28] has recently been published in this journal and proposed a new automated system of motor function post-stroke. This system utilised a single mobile phone to provide classified clinical ratings scores and was proposed as a low-cost home diagnostic device. In contrast, the goal of present paper is to develop a more comprehensive system designed for the clinical environment which can quantify hand/wrist movements and provide a score which goes beyond the standard clinical rating scores. The system is composed of wearable inertial sensing and MMG sensors. The output of the system combines classified FMA-UE scores with a series of fine-grained supplementary features which have been formulated as a means of complementing the low resolution clinical rating scores. These supplementary features are extracted directly from the sensor data and provide fine-grained information about the sub-components of motor function. Features span information about the joint range of motion and muscle performance.

II. Methods

A. Subjects and Design

Following ethical approval and informed consent, 64 subjects were recruited from the acute and hyper-acute wards at Charing Cross Hospital (London, UK). These subjects presented with the following characteristics:

- Sex: 33 male, 31 female
- Affected side: 43 left, 21 right
- Age: 66.3 ± 13.8 years old (range 33-95 years old)
- FMA-UE score: 42.7 ± 18.1 (range 0-60)
- Days since stroke: 9.3 ± 15.3 (range 2-90 days)

Upper extremity motor function evaluation was performed in the acute or subacute phases of stroke using the motor function section of the Fugl Meyer Assessment (FMA-UE) [6]. This scale is well validated as a sensitive measure of the various subcomponents and overall upper extremity function [5]. Assessments were performed by two experienced examiners (one to oversee the sensor system and one to rate the clinical score).

The study imposed no subject requirements in terms of minimum level of required motor function and a total of 9 subjects with a total FMA-UE score of ≤ 15 were included in the cohort. This is worth highlighting since the proposed system was able to test all functional impairment levels unlike many prior developed automated systems. A minimum level of cognitive ability was required and judged to be sufficient if the subject scored a full Glasgow Coma Scale Rating. This ensured the subject was in a position to be able to follow simple verbal instructions. This study received ethical approval (No.11/LO/0941) from the Health Research Authority of the

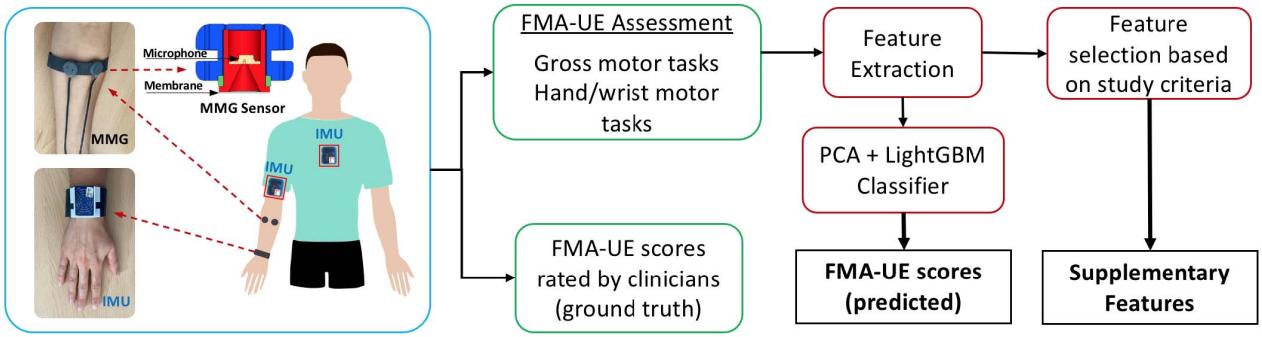

Fig. 1. Schematic diagram illustrating the instrumentation and major stages of the study. IMUs are located in three locations: torso, upper arm, wrist. MMGs are located on the underside of the forearm. Major stages of the study highlighted are the classification of FMA-UE task scores and production of the supplementary features.

National Health Service (NHS), UK. All subjects gave their full consents for participation prior to testing.

B. Data System

A collection of in-house data logger boards was used to stream all the sensor data collected from the subject. The boards were equipped with an inertial measurement unit (IMU) sensor (STMicroelectronics, Switzerland) which records tri-axial accelerometer, gyroscope, and magnetometer data. In addition, the board contained a set of ADC pins, which enabled recording from peripheral devices, and a Bluetooth module for online data transmission. These boards have been previously validated for the computation of orientation [29] and have been implemented in multiple large clinical studies [21], [22]. Housing cases for the data logger board were 3D printed in a specific shape to expose the peripheral ADC pins and to enable the attachment of 3D printed flexible resin straps for fitting to the subject (see Fig. 1).

A novel inclusion for an automated system of motor function post-stroke is the addition of MMG sensors to measure myographic data. These sensors have been developed in-house using a condenser microphone-based design and with a chamber which has been selected to maximise the frequency response [30]. The chamber is in turn sealed with an air-tight membrane to create a constant pressure and to prevent interference from external sounds. MMG sensor data was collected and synchronised with the data logger boards by sampling directly via the exposed ADC pins. The time series response of the MMG broadly resembles that of EMG except that the measurand is the vibration of the muscle rather than the electrical activity. An illustration of how this filtered response appears at different motor functional levels is displayed in Fig. 2. This data was recorded for the mass flexion task, which involves the subject transitioning from a state of full finger extension to flexion as quickly as possible.

Data were sent via Bluetooth to a tablet which was running a bespoke GUI for logging data. This software was designed in C# using the .NET framework. Functionality was implemented to enable connection to the data logger boards, online data visualisation, and data collection. Streamed data was saved onto the tablet during the testing phase and then later transferred to a secure computer for further data processing.

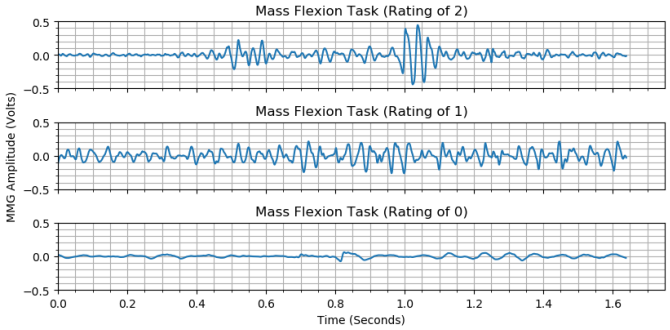

Fig. 2. Plots of filtered MMG response recorded during the mass flexion task of the FMA-UE. Each plot is taken from a different subject who scored a different rating for this task (0, 1, and 2).

C. Data Collection

Inertial data was collected from the IMU on board each of the data logger boards. These boards were attached to the body sections of the torso, upper arm, and lower arm. A lack of relevant literature meant that MMG placement was informed based on information produced for EMG instead [31]. MMGs were placed to capture muscle activity of the extrinsic finger flexors (flexor digitorum profundus and flexor digitorum superficialis) and wrist flexors (flexor carpi radialis). These muscle groups were selected as a means of quantifying finger and wrist flexion. A diagram of the MMG sensors approximately placed to capture these muscle groups is shown in Fig. 1.

Sensor data were collected during each of the tasks in the motor function section of the FMA-UE. All of the 21 motor tasks of this section were performed as part of this study. The reflex tasks were not included since these could not be measuring using the sensor system. For this study it was convenient to categorise tasks of the FMA-UE either as gross motor tasks (those involving the upper or lower arm segments) or hand/ wrist tasks (those tasks only involving the hand and/or wrist). This is because gross motor tasks could be captured using the inertial sensors whereas hand/wrist motor tasks could only be captured using the MMG data.

D. Data Processing

Data processing was performed using a custom GUI developed using the Qt framework in the Python environment. Data types calculated for this study may be categorised as

TABLE I

FEATURES EXTRACTED FROM THE THREE DATA TYPES CALCULATED USING THE NOVEL SYSTEM. INERTIAL FEATURES ARE THOSE WHICH ARE CALCULATED FROM THE ACCEROMETER OR GYOSCOPE DATA. ORIENTATION FEATURES ARE CALCULATED FROM THE ORIENATION DATA (JOINT ANGLE OR PLANE). FEATURES WHICH WERE SELECTED TO KEEP IN THEIR PRE-EXISTING FORM AS SUPPLEMENTARY FEATURES ARE HIGHLIGHTED IN RED

Inertial features		**MMG features**		**Orientation features**
Time-series	**Frequency-series**	**Time-series**	**Frequency-series**	**Time-series**
Minimum	Dominant frequency	Modified mean abs. value	Dominant frequency	Minimum
Maximum	Mean frequency	Log detector	Mean frequency	Maximum
Peak to peak	Median frequency	Average amplitude change	Median frequency	Peak to peak
Mean absolute value	Mean power	Difference abs. standard deviation	Mean power	Median crossing
Standard deviation	Frequency ratio	Zero crossing (%)	Frequency ratio	Standard deviation
Interquartile range	Peak frequency ratio	Myopulse rate (%)	Peak frequency ratio	
Median crossing		Willison amplitude (%)	Variance central frequency	
Skewness		Slope sign change (%)		
Kurtosis		Hamming window (%)		
Root mean square		Root mean square		
Power		Power		
Trapezium Integration		Trapezium Integration		

inertial, MMG, or orientation-based. Inertial and MMG data was digitally filtered prior to feature calculation using zero-phase IIR Butterworth filters.

Inertial data processing involved low pass filtering of gyroscope data at 10 Hz and band-pass filtering of accelerometer data at 1-10 Hz. MMG data was bandpass filtered at 5-100 Hz. This bandwidth has previously been shown to capture most of the relevant signal while removing the high frequency noise and low frequency motion artefacts present [32].

Orientation data, as represented by quaternions, was calculated from the inertial data collected at each instrumented body segment. Quaternions were calculated by passing the raw inertial data through the Madgwick gradient descent algorithm [33]. This algorithm has been proposed as an equally accurate and significantly less computationally intensive method than the Kalman Filter [33], which is more commonly used for this application. The orientation metrics of joint angles and plane of motion for the shoulder and elbow joints were subsequently calculated using these quaternions and the swing-twist decomposition method [34]. An avatar was produced to visualise these metrics prior to implementing them in the study.

E. Feature Extraction

Time and frequency inertial features were selected from those which have been shown to be effective in previous automated wearable systems of motor function [10], [19]. MMG features were largely selected from the list proposed by Phinyomark *et al* [35] for EMG signal classification. A collection of all the time and frequency series features calculated for each of the three data types extracted from the sensor system is displayed in Table I. The equations for the less commonly used features are displayed in full for clarity.

A subset of the time-series features, calculated for each data point (i) of the time-series vector (x) over a total length N, for the present study is as follows:

Skew: A measure of the skewness of a distribution of sensor data. Used as an inertial feature in the present study.

Normalised Median Crossing: The number of times the signal crosses the median line. Used as an inertial feature in the present study.

Trapezoidal Rule: A measure of approximating the definite integral. Used as an inertial and MMG feature in the present study.

Log Detector: Provides an estimate of the muscle contraction force. Used as an MMG feature, defined as:

$$LOG = e^{\frac{1}{N}\sum_{i=1}^{N} log(|x_i|)} \tag{1}$$

Myopulse Percentage Rate: A calculation of the overall amount of time the myopulse output exceeds a minimum threshold. Threshold value set by the Root Mean Square (RMS) of the total measured signal for the present study. Used as an MMG feature, defined as:

$$MYOP = \frac{1}{N}\sum_{i=1}^{N}[f(x_i)] \tag{2}$$
$$\text{with } f(x) = \left\{ \begin{array}{l} 1, \text{ if } x \geq RMS \\ 0, \text{ otherwise} \end{array} \right\}$$

Slope Sign Change Percentage: Provides frequency information by measuring the signal sign changes. A threshold value is set to remove false positives due to signal noise. Used as an MMG feature, defined as:

$$SSCP = \frac{1}{N-2}\sum_{i=2}^{N-1}[f[(x_i - x_{i-1})(x_i - x_{i+1})]] \tag{3}$$
$$\text{with } f(x) = \left\{ \begin{array}{l} 1, \text{ if } x \geq 0.1RMS \\ 0, \text{ otherwise} \end{array} \right\}$$

The Fourier Transform was used to transform signals to the frequency-domain, with frequency vector (f) and corresponding amplitude vector (A). Frequency is defined at each bin (j) over a total length M. A subset of the frequency-series features for the present study is as follows:

Mean Frequency: An assessment of the centre of the distribution of power across frequencies. Used as an inertial and MMG feature in the present study.

Mean Power: The mean power of the frequency transformed signal. Used as an inertial and MMG feature in the present study.

Power Ratio: The ratio of the power below (PB) and power above (PA) the mean frequency value. Used as an inertial and MMG feature, defined as:

$$PR = \frac{\sum_{j=1}^{M} PB(A(j), f(j))}{\sum_{j=1}^{M} PA(A(j), f(j))} \quad (4)$$

$$\text{where } PA(A, f) = \left\{ \begin{matrix} A^2, & \text{if } f > MNF \\ 0, & \text{if } f \leq MNF \end{matrix} \right\},$$

$$PB(A, f) = \left\{ \begin{matrix} 0, & \text{if } f \geq MNF \\ A^2, & \text{if } f < MNF \end{matrix} \right\}$$

Power Spectrum Ratio: Calculates the ratio of the power in a frequency window (WP) around the dominant frequency, and the power of the rest of the signal. The lower boundary (LB) and upper boundary (UB) define the boundaries of the window. Used an inertial and MMG feature, defined as:

$$PSR = \frac{WP}{\sum_{j=1}^{M} f(j)} \quad (5)$$

$$\text{where } WP = \sum_{j=LB}^{UB} A(j)^2$$

Variance of Central Frequency: Variance of the central frequency is an important characteristic of the frequency signal and may be defined using the spectral moments (Z). Used as an MMG feature, defined as:

$$VCF = \frac{SM[2]}{SM[0]} - \left(\frac{SM[1]}{SM[0]}\right)^2 \quad (6)$$

$$\text{where } SM[Z] = \sum_{j=1}^{M} A_j f_j^Z$$

F. FMA-UE Tasks of Interest

The FMA-UE contains a total of 21 motor tasks (not including reflex tasks) of which 9 require gross motor function and 12 belong to the hand and wrist categories. Three of the gross motor tasks ("flexor synergy", "extensor synergy", and "coordination/speed") involve multiple sub-components to score. To ensure consistency of labelling with the other tasks in subsequent classification of score, the overall scores of these tasks were downsampled to match the standard ratings assigned to the other tasks (0, 1, or 2).

A total of 7 out of the 21 tasks were excluded from the present study upon further analysis. The "mass extension" and "thumb adduction" tasks were excluded since the MMG sensors used in this study only captured hand/wrist flexion and not finger extension or thumb adduction. Secondly, despite best efforts to recruit a cohort of subjects with a wide range of deficits, some tasks were found to be assigned the maximum score (2) for over 70% of the tested subjects. These tasks were removed to avoid biasing the overall classification results. Tasks removed were "mass flexion", "hook grasp", "pincer grasp", "cylinder grasp", and "spherical grasp".

G. Classification Pipeline

A classification pipeline was developed to predict 14 of the motor tasks contained in the FMA-UE. The features for the classifiers were calculated using the equations presented in Section II-E and labels were provided by the ratings assigned during clinical testing. Only the MMG features were used to classify the hand/wrist tasks of the FMA-UE since these could not quantified by the inertial sensing in this study. The inertial-based features were used to classify all other tasks of the FMA-UE (gross motor tasks).

Dimensionality reduction was performed using principal component analysis since this method had a short computational time and was found to provide a good level of performance in the current study. The size of the feature subset was included as a hyper-parameter in the classification pipeline to ensure that the optimal number of features was preserved for each task score predicted. The optimal feature number was selected from the following set: $[3, 5, 10, 15]$.

The classification model chosen for this study was the Light Gradient Boosting Model (LightGBM) (Microsoft Corporation). This is a boosting type of classification model which uses a novel method of growing the trees for a series of decision tree classifiers. This classifier has shown great potential for speed and high performance on small datasets but has not yet been applied in an automated system of motor function. Several of the parameters of the LightGBM model were included as hyper-parameters in the classification pipeline.

The developed classification pipeline encompassed all the aforementioned classification stages including normalisation, resampling, feature reduction, and classification. Classification performance was assessed by 10 iterations of stratified 10-fold cross validation. Folds were split with respect to each subject to ensure that single subject data was not present in both the train and test set. Within each internal training fold, hyper-parameter optimisation was performed by exhaustive grid search using 5-fold cross-validation. This procedure ensured that there was no possible data leak between the training and test set at any point during classification.

H. Supplementary Features

Supplementary features were designed to provide a more fine-grained and sensitive measure of motor function than possible with the predicted clinical rating score. Supplementary features were selected from the total set of extracted features (see Section II-E) based upon the criteria of innate usefulness and transparency as clinical measures of motor function.

An objective measure of the usefulness of the feature set assessed by finding the Pearson's correlation between each

TABLE II

SUPPLEMENTARY FEATURES SELECTED FROM THE TOTAL SET OF FEATURES EXTRACTED BY THE NOVEL SYSTEM. FEATURES ARE SORTED BY THE METRIC TYPE (LOCATION AND MEASURAND), FEATURE TYPE, AND FMA-UE TASK FROM WHICH THEY WERE EXTRACTED. PEARSON CORRELATION COEFFICIENTS CALCULATED BETWEEN THE SUPPLEMENTARY FEATURES AND THE **LOCAL** (FMA-UE TASK THE FEATURE WAS EXTRACTED FROM) AND THE **GLOBAL** (SUM OF ALL FMA-UE MOTOR TASKS) SCORES

Metric	Feature	FMA-UE task	Corr. with FMA-UE (local)	Corr. with FMA-UE (global)
Shoulder twist angle	Range	Pronation-supination (elb. at 90)	0.47	0.56
Shoulder abduction angle	Maximum	Shoulder abduction	0.66	0.64
Shoulder flexion angle	Maximum	Shoulder flexion (90-180)	0.85	0.78
Elbow twist angle	Range	Pronation-supination (elb. at 90)	0.77	0.66
Elbow flexion angle	Range	Extensor synergy	0.78	0.73
Wrist flex-ext muscle activity	Difference abs. standard deviation	Rep. dorsi-volar flexion (elb. 90)	0.52	0.56
Wrist circumduction muscle activity	Difference abs. standard deviation	Circumduction	0.68	0.61

TABLE III

CLASSIFICATION PERFORMANCE RESULTS FOR THE GROSS MOTOR TASKS OF THE FMA-UE AS CLASSIFIED USING THE 3-CLASS MODEL. F1-SCORES ARE DISPLAYED FOR EACH LABEL (0, 1, 2) OF THE 3-CLASS CLASSIFICATION MODEL

FMA-UE task	F1-score (0)	F1-score (1)	F1-score (2)	Accuracy (3 class)
Flexor synergy	0.8	0.17	0.86	0.75
Extensor synergy	0.61	0.26	0.9	0.77
Hand to lumbar spine	0.64	0.55	0.83	0.72
Shoulder flexion (0-90)	0.74	0.65	0.9	0.81
Shoulder abduction	0.62	0.68	0.91	0.81
Shoulder flexion (90-180)	0.82	0.24	0.8	0.71
Pronation-supination (elb. at 90)	0.77	0.5	0.86	0.78
Pronation-supination (elb. 0)	0.51	0.24	0.85	0.69
Coordination-speed	0.71	0.54	0.82	0.71

feature and a gold-standard measure of clinical motor function. For this study, the overall FMA-UE (sum of all 21 motor tasks) was chosen as the gold-standard measure since we already recorded this measure as part of the classification pipeline. Only features which ranked amongst the top 20 correlation values were considered for selection as supplementary features.

The measure of feature transparency is more subjective but also important to consider. The supplementary features are intended to provide additional information to the clinician and as such the features must be easily understandable and not represent "black box" measures. From the top 20 most useful features, 7 features were selected which were considered to be transparent measures of motor function.

A total of 7 supplementary features remained after filtering using the aforementioned selection criteria. These features encompass aspects of joint orientation and muscle activity. These features are displayed alongside the Pearson's correlation values they achieved with the FMA-UE in Table II.

III. RESULTS

A. Classification Performance

A persistent limitation in many prior automated system studies is that the performance of the classification model was only presented as the accuracy of the predictions. This metric provides limited information as to how well each of the classes are predicted and as a consequence is not a good measure of classification performance in isolation. This problem is mitigated in the present study by displaying both the accuracy and the F1-score [36] (for each class) of the classification model.

Classification results are separated into categories for tasks involving either gross or hand/wrist motor function. All 15 of the tasks identified in Section II-E were classified using the classification pipeline discussed in Section II-G. The results for the gross motor tasks and the hand/wrist motor tasks are shown in Table III and Table IV respectively.

B. Feature Contributions

As previously discussed in Section II-G, a reduced feature set was found for each task prior to classification using PCA, and this process was repeated for each cross-validation fold. One limitation of this method is that it makes it difficult to identify the relative contributions of different features to classification. An alternative method of identifying these contributions is implemented in this study by examining the correlations between the features and clinical rating scores.

TABLE IV
CLASSIFICATION PERFORMANCE RESULTS FOR THE HAND/WRIST MOTOR TASKS OF THE FMA-UE AS CLASSIFIED USING THE 3-CLASS MODEL. F1-SCORES ARE DISPLAYED FOR EACH LABEL (0, 1, 2) OF THE 3-CLASS CLASSIFICATION MODEL

FMA-UE task	F1-score (0)	F1-score (1)	F1-score (2)	Accuracy (3-class)
Stab. at 15 dorsiflexion (elb. 90)	0.35	0.39	0.75	0.62
Rep. dorsi-volar flexion (elb. 90)	0.49	0.27	0.82	0.7
Stab. at 15 dorsiflexion (elb. 0)	0.46	0.41	0.77	0.65
Rep. dorsi-volar flexion (elb. 0)	0.32	0.18	0.75	0.55
Circumduction	0.6	0.42	0.69	0.58

TABLE V
FEATURE CONTRIBUTIONS AS ASSESSED ACROSS THE GROUPS OF **GROSS** AND **HAND/WRIST** MOTOR TASKS. FEATURE COMPONENTS (MODALITY, LOCATION, AND FEATURE TYPE) ARE RANKED BASED ON THE FREQUENCY WITH WHICH THEY OCCURRED IN THE SUBSET OF TOP 5 MOST IMPORTANT FEATURES ACROSS ALL MOTOR TASKS WITHIN THE GROUP. THE HIGHEST CONTRIBUTING MODALITY AND LOCATION IS NOT SHOWN FOR THE HAND/WRIST GROUP SINCE THESE TASKS ARE ONLY CAPTURED BY ONE MODALITY (MMG) IN ONE LOCATION (LOWER ARM)

Rank	Modality (gross)	Location (gross)	Feature type (gross)	Feature type (hand/wrist)
1	Gyroscope	Lower Arm	Standard Deviation	Average change in amplitude
2	Joint angle	Shoulder	Maximum	Difference in abs. standard deviation
3	Joint plane	Upper arm	Peak to peak	Log detector
4	Accelerometer	Elbow	Interquartile range	Modified mean absolute value
5		Torso	Root mean square	Hamming window

For each task, the Spearman's rank correlation coefficient was found between the feature value and clinical rating score across all subjects tested. Next, the features were ranked based on this coefficient to identify the five features which showed strongest correlation with the clinical rating score. Since it would not be possible to display these results for every single task classified, the features which ranked most in the top five most commonly across the different groups of tasks (gross or hand/wrist) are presented in Table V.

C. Supplementary Features Concurrent Validity

The features selected to supplement the classification score had to demonstrate their usefulness as measures of motor function. The most feasible method of achieving this is by calculating the concurrent validity with a pre-existing well established measure. The current gold-standard measures of motor function post-stroke are the standard clinical rating scales and of these the FMA is one of the most widely used and validated. Pearson correlation coefficients were calculated between each of the supplementary features and the task (local) and overall (global) FMA-UE scores as shown in Table II.

IV. DISCUSSION

The present study sought to build on prior automated systems of upper extremity motor function post-stroke in two principal ways: (1) the integration of MMG sensing to enable classification of hand/wrist motor tasks and (2) develop a model which combines a predicted output of standard clinical rating score with a series of fine-grained features. These supplementary features are calculated from sensor-derived metrics of the subcomponents of motor function and provide a continuous measure of motor function as opposed to the broad discrete values offered by the clinical rating scores. This section will be discussed in terms of meeting these two objectives and how the classification performance compares to prior automated studies.

A. Classification of the FMA-UE

Gross motor tasks of the FMA-UE were classified with a mean accuracy of 75%. There is also a relatively low range of accuracy with a minimum value of 71% and maximum of 81% showing high classifier precision. Tasks which were less well classified were those which involved more complex or several stages of movements including the "Coordination-speed" (71%), "Flexor synergy" (75%), and "Hand to lumbar spine" (72%) tasks. All clinical assessments were performed at the bedside (in many cases only a few days post-stroke) and this likely had a negatively effect on the classification accuracy of all the tasks. The decision to test at the bedside was made so that the results would a better representation of how the system would perform in clinical practice. This approach (compared to having the subject seated at a table for instance) meant that the subjects had no consistent body position and as such performed the tasks in less predictable ways. The lack of consistency within each task in turn makes it more difficult for a classification model to predict motor impairment level.

The hand/wrist motor tasks, which were classified using the MMG-derived features alone, scored a lower overall classification accuracy of 62% as compared to the gross motor tasks. This drop in performance may be due to the small movements involved in these tasks which are more difficult to quantify than gross motor tasks. Another reason may be that the MMG-derived features are less sensitive than inertial for the rating

TABLE VI
SUMMARY OF THE PROTOCOL AND RESULTS OF THE PRESENT AND RELEVANT PAST STUDIES OF AUTOMATED SYSTEMS OF MOTOR FUNCTION POST-STROKE. CLASSIFICATION RESULTS ARE FOR TASKS OF THE FMA-UE WHICH ARE CATEGORISED EITHER AS **GROSS** OR **H/W** (HAND/WRIST) MOTOR TASKS IN THE PRESENT STUDY. ONE STUDY USED TASKS FROM THE MALLET SYSTEM AND THESE TASKS HAVE ALL BEEN CATEGORISED AS GROSS MOTOR TASKS IN THIS TABLE. NA (NOT AVAILABLE) VALUES ARE MARKED WHERE THESE VALUES WERE NOT PROVIDED BY THE AUTHORS

Authors	Sensor system	Testing conditions	Task set	Stroke subjects	Gross class. accuracy (%)	H/W class. accuracy (%)
Kim et al (2016)	Kinect camera	Subject seated in instrumented room	6 FMA-UE tasks	41	Range 65-87 (mean=NA)	Not performed
Lee et al (2018)	Kinect camera + force sensing	Subject seated in instrumented room	19 FMA-UE tasks	9	Range 67-100 (mean=93)	Range 67-100 (mean=91)
Seo et al (2019)	Kinect camera	Subject seated in instrumented room	5 Mallet system tasks	7	Range 43-100 (mean=77)	Not performed
Present Study (2021)	IMU + MMG	At the bedside	14 FMA-UE tasks	64	Range 69-81 (mean=75)	Range 55-67 (mean=63)

of motor function. The application of MMG sensors for this application is still in its infancy and as such it is likely that more powerful features for classifying MMG data will be proposed in the future. Overall, these early results suggest the promise of a modality which may be applied more easily than inertial sensing and does not suffer from the skin impedance and robustness concerns that are associated with EMG.

Evaluation of the F1-scores for the three-class classification model developed for the present study is apparent that the model has difficulty predicting the mid-rating score (label 1). This is evidenced by a mean F1-score of 0.44 and 0.34 for the gross motor and hand/wrist motor tasks respectively. One reason for this may be due to noise inherent in the labelling. The ratings assigned to the FMA-UE are inherently broad and can incur a deal of ambiguity as to what constitutes partial movement as compared to full or no movement. Another contributing factor to a low F1-score for this label may be the high inter-variability in how subjects perform tasks at this functional level. A subject with only partial motor function may require a variety of different compensatory movements to achieve a given task, which may make classification more difficult. This is in contrast to a subject with no or full motor function, who would be expected to perform the task in a more predictable manner. A final reason for the low F1-score is an imbalanced dataset due to this label not being well represented in the data. The best solution for the final two limitations is to collect additional training data so that the classifier is better equipped to classify these labels.

There are a limited number of comparable high quality automated studies of upper-extremity motor function post-stroke with which to compare the results of the present system. A number of prior systems were applied in pilot studies and therefore the classification results achieved have limited validity. Many other automated studies were proof-of-concept and only sought to find correlations between features or custom scores with clinical rating scales. Several studies used a bilateral measure whereby a comparison between a paretic and non-paretic arm was calculated, and therefore results may not be compared to the present study. Finally, several of the remaining studies lacked a rigorous enough classification procedure or utilised too small a sample size. Three prior studies were identified as being suitable for comparison with the present studies and these are displayed in Table VI. The main differences between the past and the present study is the instrumentation used and the required testing conditions.

Kim *et al* [37] developed and trialled the application of an automated system (single Kinect depth-sensing camera) of the FMA-UE in a large (41 hemiplegic stroke subjects) clinical study. A classification model was developed to classify 6 of the tasks from the FMA-UE. The remaining tasks of the FMA-UE were not incorporated, presumably due to difficulty detecting these movements with the Kinect sensor, such as those involving twist rotation around the bone axis or hand/wrist motor movements. The study by Kim *et al* achieved prediction accuracies for each task ranging from 65% to 87%. These results are comparable to the present study which ranged between 71% and 81% for the gross motor tasks, which cover a similar task set.

The study by Lee *et al* [7] extended the sensor system proposed by Kim *et al* with the inclusion of force sensing resistor in addition to the Kinect sensor. This enabled the hand grasp tasks to also be quantified. The study by Lee *et al* tested subjects in an instrumented room complete with Kinect sensor and instrumented tools. Clinical scores were predicted using a rule-based classifier with features extracted based on the guidelines of the FMA-UE. This meant that the performance of the classification model was not limited by the relatively small sample size (9 subjects) recruited for the study. Overall, the study achieved a very high classification accuracy of approximately 92% over all tasks tested. One caveat to the work performed by Lee *et al* is that testing was performed in a very controlled environment which means that similar results may not be achieved in a normal clinical environment. The system produced in the present study was not able to achieve comparably high classification accuracy but does offer several advantages. For instance the instruments applied in the present study may be used to test subjects at all levels of motor function and may be tested at the subjects bedside, which increases their ease of use in a clinical setting. The results of the present study were also achieved

over a much larger cohort size which increases their clinical significance.

A final clinical study has been performed to evaluated the automated system developed by Seo *et al* [13]. The proposed system (similar to the prior two studies) implemented the Kinect sensor for motion tracking. Unlike the study by Lee *et al* which implemented a secondary instrument to compensate for movements which the Kinect sensor cannot detect, the study by Seo *et al* instead chose to classify the tasks of the Mallet clinical rating scale, which does not require these movement types. A rule-based classification model was developed for predictions to avoid the requirement for a significant amount training data (only 7 subjects tested). The accuracy scores achieved for these tasks ranged from 43% to 100% and averaged 77%. The mean result is comparable to those achieved in the present study (the range is much larger). The Mallet clinical scale, although well suited for measurement using the Kinect sensor, is not traditionally used in stroke evaluation and as such may have limited clinical validity.

B. Supplementary Features

All supplementary features selected (see Table II) exhibited a positive Pearson correlation coefficient with the overall (range 0.56 - 0.78) and task-specific (range 0.47 - 0.85) FMA-UE scores, which were used as the gold-standard measures.

The study by Fu *et al* [38] investigated the concurrent validity of a shortened FMA with other clinical rating scales using the Pearson correlation coefficient. This study considered a correlation coefficient of >0.75 to be excellent, 0.5-0.75 to be good, and 0.25-0.5 to be fair. The shortened FMA was found to have a correlation of 0.57 (good) with the Stroke Impact Scale hand function subscale. On this basis, the correlation coefficients measured between the supplementary features and the overall clinical score in the present study all range in the good to excellent range (0.56 - 0.78). This suggests high concurrent validity of the extracted features with the FMA-UE and supports their usefulness as supplementary metrics of motor function post-stroke.

The study by Julianjatsono *et al* [15] developed a sensor system consisting of the Kinect sensor and a wearable glove. This system was utilised in the development of a regression model to predict the outcome of six tasks of the FMA-UE. A unique feature was calculated to define each of the six tasks, and the Pearson correlation coefficient calculated with the clinical score assigned for that task. Correlation coefficients were achieved which ranged from 0.17 to 0.475. This is compared to the present study which found correlations of 0.47 to 0.85 using the novel features developed over a similar task set (only one of the tasks selected differs between the studies). These results give a promising indication of the usefulness of the features extracted in the present study as measures of motor function as compared to the wider literature.

C. Clinical Implications

Existing automated systems of motor function post-stroke have variously been proposed as home-based rehabilitative aids, or as replacements for clinical testing in monitoring subject improvement or research applications. The system developed in the present study was designed to ensure the testing process was as seamless as possible for both clinician and subject alike. Unlike many prior automated systems, assessment could be performed at the bed-side and was possible even for bed-bound subjects. In addition, there was no requirements for remote cameras to be set up or for the subjects to be in a sitting position. Despite this design, the administration of the system still presented an increased time, cognitive, and physical burden on clinician and subject alike due to the attachment of wearable sensors, and extra consideration to make sure each task was recorded correctly. These drawbacks mean that automated systems are unlikely to be widely adopted in the near future as means of monitoring subject improvement as part of normal post-stroke care.

In the opinion of the authors, the better application for automated systems is in a field whereby more time may be afforded, and the high resolution provided by the fine-grained supplementary feature set may lead to useful insights. The developed automated system could be used to enrich studies of medicinal or rehabilitative interventions by improving the set of motor outcomes. The study presented in this paper illustrates the potential of an automated system which derives sensor-based metrics providing insights beyond simply predicting standard clinical rating scales.

V. Conclusion

The first novel contribution of this paper is the development of a system of motor function post-stroke which combines a classified clinical rating score with several sensor-derived supplementary features. This combination of metrics incorporates the validation of a well-known clinical score with the high resolution provided by the motor features. Classification performance was found to be broadly comparable to the wider literature which testing a much less controlled environment. The supplementary features were shown to provide useful measures of motor impairment. These features could be used to provide information beyond the standard clinical rating scales.

The second major contribution of this paper is the introduction of MMG sensing in wearable motor system post-stroke. MMG sensing was proposed as a method to determine the motor function of the wrist and hand, a region which has either been ignored or classified using less practical sensor systems in prior systems. MMG was successfully implemented in the present study to classify the hand/wrist tasks of the FMA-UE and to derive supplementary features of muscle activity. It is expected that performance could be further improved in the future by instrumenting more regions and developing more sophisticated motor features. In particular, the MMG system should be expanded to also capture the muscle groups active in finger/wrist extension and thumb adduction since these movements form a major role in several tasks of the FMA-UE

Finally, a persistent limitation of prior automated systems of motor function post-stroke is that validation studies are conducted with low number of subjects and in very controlled circumstances. This means that any results of the systems

are unlikely to be replicated in real clinical environments. This was rectified in the present study by testing the system over a large subject cohort, and assessing the subject directly at the bedside. The sensing system has been successfully patented [39] with translational activity underway for widespread in-clinic and home-based use.

ACKNOWLEDGMENT

The authors express their sincerest gratitude our test subjects at Charing Cross Hospital and the support of Serg Technologies, the U.K. NIHR BRC, the U.K. UKRI EPSRC Centre for Doctoral Training in Neurotechnology for Life and Health, the U.K. EPSRC, and the U.K. DRI-CRT.

REFERENCES

[1] C. O. Johnson *et al.*, "Global, regional, and national burden of stroke, 1990–2016: A systematic analysis for the global burden of disease study 2016," *Lancet Neurol.*, vol. 18, no. 5, pp. 439–458, 2019.

[2] R. W. Teasell, N. C. Foley, S. K. Bhogal, R. Chakravertty, and A. Bluvol, "A rehabilitation program for patients recovering from severe stroke," *Can. J. Neurol. Sci.*, vol. 32, no. 4, pp. 512–517, May 2005.

[3] N. Rösser and A. Flöel, "Pharmacological enhancement of motor recovery in subacute and chronic stroke," *Neurorehabilitation*, vol. 23, no. 1, pp. 95–103, Mar. 2008.

[4] S. Handelzalts, I. Melzer, and N. Soroker, "Analysis of brain lesion impact on balance and gait following stroke," *Frontiers Hum. Neurosci.*, vol. 13, pp. 1–9, May 2019.

[5] M. H. Rabadi and F. M. Rabadi, "Comparison of the action research arm test and the Fugl-Meyer assessment as measures of upper-extremity motor weakness after stroke," *Arch. Phys. Med. Rehabil.*, vol. 87, no. 7, pp. 962–966, 2006.

[6] A. R. Fugl-Meyer, L. Jääskö, I. Leyman, S. Olsson, and S. Steglind, "The post-stroke hemiplegic patient. 1. A method for evaluation of physical performance," *Scand. J. Rehabil. Med.*, vol. 7, no. 1, pp. 13–31, 1975.

[7] S. Lee, Y.-S. Lee, and J. Kim, "Automated evaluation of upper-limb motor function impairment using Fugl-Meyer assessment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 1, pp. 125–134, Jan. 2018.

[8] B. Dehbandi *et al.*, "Using data from the Microsoft Kinect 2 to quantify upper limb behavior: A feasibility study," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 5, pp. 1386–1392, Sep. 2017.

[9] A. Scano, A. Chiavenna, M. Malosio, L. M. Tosatti, and F. Molteni, "Kinect V2 implementation and testing of the reaching performance scale for motor evaluation of patients with neurological impairment," *Med. Eng. Phys.*, vol. 56, pp. 54–58, Jun. 2018.

[10] S. Huang *et al.*, "Motor impairment evaluation for upper limb in stroke patients on the basis of a microsensor," *Int. J. Rehabil. Res.*, vol. 35, no. 2, pp. 161–169, 2012.

[11] V. T. Cruz, V. F. Bento, D. D. Ribeiro, I. Araújo, C. A. Branco, and P. Coutinho, "A novel system for automatic classification of upper limb motor function after stroke: An exploratory study," *Med. Eng. Phys.*, vol. 36, no. 12, pp. 1704–1710, 2014.

[12] H. T. Jung, H. Kim, M. Y. Oh, T. Ryu, and Y. Kim, "Learning classifier to evaluate movement quality in unassisted pick-and-place exercises for post-stroke patients: A preliminary study," in *Proc. IEEE Eng. Med. Biol. Soc.*, Jul. 2017, pp. 2490–2493.

[13] N. J. Seo *et al.*, "Capturing upper limb gross motor categories using the Kinect sensor," *Amer. J. Occupat. Therapy*, vol. 73, no. 4, pp. 1–10, 2019.

[14] P. Otten, J. Kim, and S. Son, "A framework to automate assessment of upper-limb motor function impairment: A feasibility study," *Sensors*, vol. 15, no. 8, pp. 20097–20114, Aug. 2015.

[15] R. Julianjatsono, R. Ferdiana, and R. Hartanto, "High-resolution automated Fugl-Meyer assessment using sensor data and regression model," in *Proc. 3rd Int. Conf. Sci. Technol.-Comput. (ICST)*, Jul. 2017, pp. 28–32.

[16] X. Hu, A. K. Suresh, W. Z. Rymer, and N. L. Suresh, "Assessing altered motor unit recruitment patterns in paretic muscles of stroke survivors using surface electromyography," *J. Neural Eng.*, vol. 12, no. 6, Dec. 2015, Art. no. 066001.

[17] X. Zhang, Z. Wei, X. Ren, X. Gao, X. Chen, and P. Zhou, "Complex neuromuscular changes post-stroke revealed by clustering index analysis of surface electromyogram," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 2105–2112, Nov. 2017.

[18] E. Repnik, U. Puh, N. Goljar, M. Munih, and M. Mihelj, "Using inertial measurement units and electromyography to quantify movement during action research arm test execution," *Sensors*, vol. 18, no. 9, pp. 1–23, 2018.

[19] Y. Li, X. Zhang, Y. Gong, Y. Cheng, X. Gao, and X. Chen, "Motor function evaluation of hemiplegic upper-extremities using data fusion from wearable inertial and surface EMG sensors," *Sensors*, vol. 17, no. 3, p. 582, Mar. 2017.

[20] R. B. Woodward, S. J. Shefelbine, and R. Vaidyanathan, "Pervasive monitoring of motion and muscle activation: Inertial and mechanomyography fusion," *IEEE/ASME Trans. Mechatronics*, vol. 22, no. 5, pp. 2022–2033, Oct. 2017.

[21] W. Huo *et al.*, "A heterogeneous sensing suite for multisymptom quantification of Parkinson's disease," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 6, pp. 1397–1406, Jun. 2020.

[22] L. Formstone, M. Pucek, S. Wilson, P. Bentley, A. McGregor, and R. Vaidyanathan, "Myographic information enables hand function classification in automated Fugl-Meyer assessment," in *Proc. 9th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, Mar. 2019, pp. 239–242.

[23] J. H. Burridge *et al.*, "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. Therapy*, vol. 41, pp. S32–S38, Jul. 2017.

[24] S. Wilson *et al.*, "Formulation of a new gradient descent MARG orientation algorithm: Case study on robot teleoperation," *Mech. Syst. Signal Process.*, vol. 130, pp. 183–200, Sep. 2019.

[25] X. L. Hu, K. Y. Tong, and L. Li, "The mechanomyography of persons after stroke during isometric voluntary contractions," *J. Electromyogr. Kinesiol.*, vol. 17, no. 4, pp. 473–483, 2007. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16603386

[26] E. V. Olesh, S. Yakovenko, and V. Gritsenko, "Automated assessment of upper extremity movement impairment due to stroke," *PLoS ONE*, vol. 9, no. 8, Aug. 2014, Art. no. e104487.

[27] M. Zhang, B. Lange, C.-Y. Chang, A. A. Sawchuk, and A. A. Rizzo, "Beyond the standard clinical rating scales: Fine-grained assessment of post-stroke motor functionality using wearable inertial sensors," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 6111–6115.

[28] X. Song, S. Chen, J. Jia, and P. B. Shull, "Cellphone-based automated Fugl-Meyer assessment to evaluate upper extremity motor function after stroke," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 2186–2195, Oct. 2019.

[29] S. O. H. Madgwick, S. Wilson, R. Turk, J. Burridge, C. Kapatos, and R. Vaidyanathan, "An extended complementary filter for full-body MARG orientation estimation," *IEEE/ASME Trans. Mechatronics*, vol. 25, no. 4, pp. 2054–2064, Aug. 2020.

[30] A. O. Posatskiy and T. Chau, "Design and evaluation of a novel microphone-based mechanomyography sensor with cylindrical and conical acoustic chambers," *Med. Eng. Phys.*, vol. 34, no. 8, pp. 1184–1190, Oct. 2012.

[31] J. Basmajian, *Biofeedback: Principles and Practice for Clinicians*, 3rd ed. Philadelphia, PA, USA: Lippincott Williams and Wilkins, 1989.

[32] C. Orizio, "Muscle sound: Bases for the introduction of a mechanomyographic signal in muscle studies," *Crit. Rev. Biomed. Eng.*, vol. 21, no. 3, pp. 201–243, 1993.

[33] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *Proc. IEEE Int. Conf. Rehabil. Robot.*, Jun. 2011, pp. 1–7.

[34] P. Dobrowolski, "Swing-twist decomposition in Clifford algebra," 2015, *arXiv:1506.05481*. [Online]. Available: https://arxiv.org/abs/1506.05481

[35] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7420–7431, Jun. 2012.

[36] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and F-score, with implication for evaluation," in *Advances in Information Retrieval*, D. E. Losada and J. M. Fernandez-Luna, Eds. Berlin, Germany: Springer, 2005, pp. 345–359.

[37] W. S. Kim, S. Cho, D. Baek, H. Bang, and N. J. Paik, "Upper extremity functional evaluation by Fugl-Meyer assessment scoring using depth-sensing camera in hemiplegic stroke patients," *PLoS ONE*, vol. 11, no. 7, pp. 1–13, 2016.

[38] T. S.-T. Fu *et al.*, "Psychometric comparison of the shortened Fugl-Meyer assessment and the streamlined wolf motor function test in stroke rehabilitation," *Clin. Rehabil.*, vol. 26, no. 11, pp. 1043–1047, Nov. 2012.

[39] R. Vaidyanathan, N. Nowlan, R. Woodward, and S. Shefelbine, "Biomechanical activity monitoring," U.S. Patent 10 335 080, Jul. 2, 2019.