IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 20 June 2024, accepted 11 July 2024, date of publication 19 July 2024, date of current version 29 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3430834

RESEARCH ARTICLE

A Cutting-Edge Approach to Decision-Making in Awake Neurosurgery Based on Hand Motion Recognition

LUIGI GABRIEL TROCONIS, (Student Member, IEEE), RICCARDO FELICETTI, (Member, IEEE), AND ANDREA MONTERIÙ, (Member, IEEE)
Department of Information Engineering, Università Politecnica delle Marche, 60131 Ancona, Italy

Corresponding author: Luigi Gabriel Troconis (l.g.troconis@pm.univpm.it)

This research has received funding from the project Vitality – Project Code ECS00000041, CUP I33C22001330007 – funded under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.5 – 'Creation and strengthening of innovation ecosystems,' construction of 'territorial leaders in R&D' – Innovation Ecosystems – Project 'Innovation, digitalization and sustainability for the diffused economy in Central Italy – VITALITY' Call for tender No. 3277 of 30/12/2021, and Concession Decree No. 0001057.23-06-2022 of Italian Ministry of University funded by the European Union – NextGenerationEU.

ABSTRACT In Awake Neurosurgery (AN), the assessment of the patient's capabilities is of paramount importance to minimize the risk of post-operative cognitive, language, and motor deficits. To retain fine hand motion capabilities, in the current clinical practice, a neuropsychologist assigns hand motion tasks to the patient during AN, and these tasks are evaluated by visual observation. This study introduces an innovative and non-invasive task evaluation method that employs an infrared stereo camera for precise hand pose acquisition. The primary contribution of this work is to offer objective, quantitative and reproducible task evaluations, addressing a critical aspect of patient care during AN procedures. The proposed method focuses on dynamic hand gestures recognition and utilizes unidirectional and bidirectional Long Short-Term Memory networks to assess common motor tasks during AN. To evaluate the tasks, we employ state-of-the-art features, both distance-based and angle-based, formed by the finger bones of the human hands. These feature vectors demonstrate a promising accuracy and inter-patient portability for the tasks under consideration, with mean accuracy exceeding 80% when tested on new patients, separate from those used in training. This approach provides an efficient solution for the identification and assessment of tasks, also eliminating the need for task-specific labeling. This, in turn, enhances usability and reduces the potential for human error. The proposed method has the potential to improve clinical decision-making in AN and offers reliable classification tools for neuropsychologists.

INDEX TERMS Awake neurosurgery, decision making support, hand gesture recognition, leap motion controller, long short term memory, recurrent neural network.

I. INTRODUCTION

Awake Neurosurgery (AN), also known as awake craniotomy, is a well-established form of neurosurgery where the patient remains awake during most of the procedure to monitor specific functions such as motor skills, vision, and cognition [1]. This technique is utilized for various applications, including the resection of brain vascular malformations, tumor removal, and treatment of lesions in language areas of the brain [2]. In current surgical practices, the assessment of a patient's functions during AN is still heavily relied on visual observation conducted by the neuropsychologist [3], thus making qualitative evaluations based on their experience. The neuropsychologist evaluates whether the patient's capabilities are preserved and guides the surgeon to minimize damage while achieving the target, such as removing cancerous tissue.

The associate editor coordinating the review of this manuscript and approving it for publication was Angel F. García-Fernández.

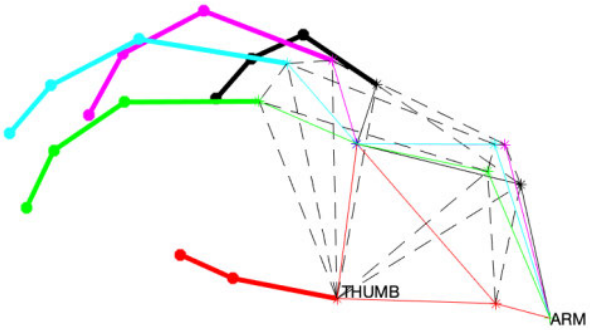

FIGURE 1. The skeletal representation, color identifies each finger: red for the thumb, green for the index finger, cyan for the middle finger, magenta for the ring finger, and black for the pinky finger.

As a result, these assessments rely on subjective judgment, playing a crucial role during AN. Functional Mapping (FM), a widely used technique in neurosurgery, aims to identify and map the brain regions responsible for language, motor skills, and cognitive functions [4]. These critical areas, known as eloquent regions, ensure proper brain function. During FM, the neurosurgeon employs targeted stimulation of specific brain regions while the patient is awake [5], [6]. This enables real-time monitoring of the patient's responses, allowing the surgeon to assess the functional organization of the brain. The identification and mapping of these eloquent areas are vital as they guide the surgical procedure, ensuring that these regions are safeguarded from damage or disruption [5], [6], [7], [8]. By minimizing the risk of post-operative cognitive, language, and motor deficits, FM contributes to improved surgical outcomes and preservation of essential brain functions [9]. In the field of AN, previous studies have explored various techniques for FM. These techniques range from invasive methods, such as cortical and subcortical electrical stimulation [7], [8], [10], to non-invasive approaches like the use of tablets [11], [12], specifically for language and cognitive mapping. Invasive techniques involving electrical stimulation provide direct insight into brain function, while non-invasive methods, such as tablet-based assessments, offer a more convenient and patient-friendly approach for functional mapping in AN [12], [13].

However, in the clinical practice, hand motion tasks are traditionally evaluated by a neuropsychologist by visual observation [3]. Automatic task recognition and evaluation could support the neuropsychologist to reduce possible human error. To the best of the authors' knowledge, specific tools for AN have never been investigated in the literature. Hence, this work aims to fill this research gap by exploring how hand motion recognition methods for dynamic hand gesture recognition can be employed to tackle this specific problem. In detail, in this paper we propose a novel approach that provides objective, quantitative and reproducible evaluations to support the neuropsychologist's decision-making in the specific case for motor mapping of hand motion tasks. While related studies on AN protocols have primarily focused on functional mapping for language and cognitive tasks [14], this research uniquely addresses the assessment of motor tasks in the context of AN. By proposing a non-invasive approach based on a infrared stereo camera, this study offers an innovative contribution to the neurosurgical field, providing insights into motor function assessment during surgical procedures. The proposed study involves the utilization of the Leap Motion Controller (LMC) infrared stereo camera for hand pose acquisition in AN. The LMC allows for high precision, with hand pose accuracy below 0.2 mm and an average of 1.2 mm in dynamic cases [15]. The data captured by the LMC are processed using skeletal-based modeling, wherein the spatial coordinates of the hand are reconstructed to create a virtual representation of a skeleton hand, as shown in Figure 1. However, given the nature of motor task assessment, we approach the problem as a dynamic hand gesture recognition problem based on feature-based system.

In this work, we focus on two specific tasks that are employed in the clinical practice to assess hand motion capabilities, namely, Open-Close (OC) and Thumb towards each of the other Fingers tips (TF). These two tasks are well-known in the AN literature. In [16], the authors evaluate hand motion tasks for functional magnetic resonance imaging, that is a non-invasive alternative to direct cortical stimulation, where the latter represents the gold standard. The authors describe two motor tasks, finger–thumb tapping and self-paced clenching and spreading of the hand, that correspond to TF and OC in this work, respectively. In [17], the authors classify hand motions from electrocorticography signals during intraoperative awake craniotomy. The analyzed hand motions are grasp (i.e., OC), thumb-finger motion (i.e., TF), and index-finger motion (a less common variant of TF).

As a result, we employ sequential models to classify the OC and TF tasks. Specifically, we utilize the Unidirectional Long Short-Term Memory (Uni-LSTM) and the BiDirectional Long Short-Term Memory (Bi-LSTM) models [18], [19], [20] to classify the performance of this two common motor task during AN. These Recurrent Neural Networks (RNNs) are applied to identify the task and determine whether it is executed correctly.

We make the following contributions:

- Proposing a contactless hand pose estimation technique for supporting the decision-making of neuropsychologists during motor-mapping procedures.
- Comparing and evaluating the feasibility of implementing Uni-LSTMs and Bi-LSTMs in a time constrained scenario.
- Identifying the combination of proposed features that provides the highest discrimination in the AN motor tasks under investigation.

The remainder of the paper is structured as follows. Section II provides an overview of the devices used in AN, the LMC application fields, and the methods for hand gesture recognition. The proposed method is detailed in Section III.

Results and discussions are presented in Sections IV and V, respectively. Finally, conclusions and future works are drawn in Section VI.

II. RELATED WORKS

In the current research on functional mapping in AN, the most widely accepted and reliable method for intraoperative brain mapping is known as Direct Cortical Electrical Stimulation (DCES) [10], [11], [12]. Currently, DCES is implemented in conjunction with various tests involving language assessment, cognitive evaluations [4], [11], [12], [21], and motor tasks [17]. In [4], the research findings reveal that all the recruited doctors (i.e., 100% of them) employ DCES as their primary method for motor mapping, and the second most frequently used technique is Electrocorticography (ECoG). In the context of motor mapping, instead, the control criterion is based on observing whether there is any noticeable motor movement during awake mapping [4].

In the study [17], the authors verified the possibility of classifying three types of hand movements (grasp, thumb-finger motion, and index-finger motion) using recorded ECoG signals during intraoperative surgeries from 4 patients (2 with brain tumors, undergoing AN, and 2 with intractable epilepsy). However, the proposed method is performed offline and requires positioning up to 95 electrodes on the brain cortex. Given that the existing literature on hand motion classification during AN predominantly relies on invasive techniques [22], we propose an alternative contactless approach using the LMC. By utilizing the LMC, we aim to achieve hand motion classification without the need for invasive procedures, for instance the ECoG use reported in [17]. This contactless method has the potential to provide accurate and reliable results while minimizing the risks and discomfort associated with invasive techniques. Our study contributes to the field by exploring the feasibility and effectiveness of using the LMC for hand motion classification, providing a contactless alternative for assessing hand movements. The LMC has demonstrated its versatility in various fields, ranging from video games [23] and augmented reality [24] to sign language recognition [19], [25], [26], [27], [28], [29]. Some studies have explored different applications of the LMC, including the assessment of motor symptoms of bradykinesia and tremor in Parkinson's disease [30], [31], [32], and its intraoperative use during deep brain stimulation to assess motor symptoms [30], [33].

In order to provide a comprehensive overview, in the following we include a discussion on the current state of research in hand gesture recognition, in particular to sign language recognition, which is a strictly related field in the literature. The existing literature on the topic can be classified into deep learning and feature-based systems. Deep learning approaches are powerful tools for problems related to big data [34], [35], [36]. However, in scenarios where a large amount of data is not required or difficult to collect, it may be advantageous to extract features from the data, namely, feature-based approaches. The hand motion task can be further categorized into static and dynamic hand motion recognition. Static gesture recognition methods achieve high accuracy for static gesture words [37], however, they cannot recognize the meaning associated with the movement of the hand itself. Thus, they are unsuitable for the purpose of this paper, where the hand motion is the primary focus. Research on dynamic hand recognition is composed of single-hand and double-hand gesture recognition [38], [39]. In this paper, we focus on the study of single-hand dynamic gesture recognition using the LMC. One of the pioneering works utilizing LMC for American Sign Language (ASL) recognition is presented in [40]. This research proposes a Multi-Layer Perceptron neural network with a backpropagation algorithm to classify 26 symbols of ASL, achieving a recognition rate of 96.15% on a test dataset. In [19], a hand gesture recognition method is proposed, employing RNNs and the LMC. The approach utilizes angles formed by finger bones and demonstrates high efficiency in recognizing a wide range of gestures, including those from ASL. A methodology utilizing the LMC and RNN has also been proposed for Arabic Sign Language recognition [26]. This method achieves an average classification accuracy of 89% for single-hand gestures and 96% for two-hand gestures. Recently, a sequential learning approach based on the spatio-temporal prosodic and angle features has been proposed as well [28]. The method utilizes 3D hand skeletal information from LMC and encodes it with a Fast Fisher Vector before processing with a Bi-LSTM. This approach achieves a recognition accuracy of 98.6% for a selected ASL dictionary and 91% for pairs of similar ASL words. The classification system [41] for activities of daily living based on grasp motion, employing the LMC, achieves 99% accuracy. The study highlights two features, namely the adjacent tips distance and joint angle, which prove to be particularly effective for classification of daily activities. The study in [42] focuses on recognizing seven fist signs in ASL utilizing the LMC to detect the position of the thumb, which is crucial for differentiating these signs. In [43], two classifiers (Support Vector Machine (SVM) and deep neural network) are compared to develop a sign language recognition prototype using the LMC. This research shows the superior performance of the deep neural network in recognizing ASL hand gestures compared to the SVM, with an improvement of 13% in terms of accuracy. In [44], an approach employing Bi-LSTM with k-Nearest Neighbour method for classifying ASL alphabets, trained on 2600 samples, is reported. This method is designed for real-time recognition and has been integrated into a game-like application for ASL learning. The recent research [29] proposes a methodology for recognizing continuous signed letters from backhand view using a rewound video sequence and previous signed-letter information. The classification employs Long Short-Term Memory (LSTM) based on time-independent patterns, considering the influence of previous letters on the current one. The approach improves the

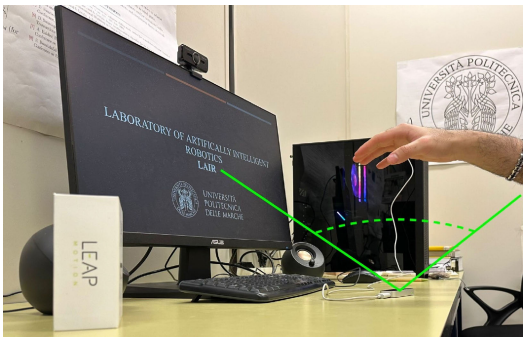

FIGURE 2. Acquisition setup.

accuracy of signed-letter recognition by incorporating the dynamics of hand motion and the context of previous gestures.

III. MATERIALS AND METHOD

In this section, we present the proposed method for classifying the OC and TF tasks. The proposed framework consists of data acquisition, preprocessing, data augmentation, feature extraction, filtering, and classification. The entire study is conducted in the MATLAB environment.

A. SUBJECTS AND DATA ACQUISITIONS

For acquiring the spatial coordinates sequence of the hands, the LMC is employed. LMC is a hand tracking system that employs stereo infrared cameras. It offers a comfortable interaction zone, up to a distance of 80 cm, and has a wide field of view (120° × 150°). Additionally, it can operate in various environmental conditions. However, the LMC has limitations, which are reported in [45], such as:

- Variable frame rate, which can vary between 20 and 200 frames per second, and it can be affected by computer settings and the USB version being used.
- Occlusions, that must be avoided as much as possible.
- Light conditions and background.

In the case of AN, the relevant limitations are the frame rate and the occlusions. In the surgery room, the light conditions are not critical if the device is fixed below the hand during the acquisition. Additionally, the background is static and almost uniform throughout the entire surgery room.

For the purpose of this study, a handcraft dataset was created. The dataset consists of recordings from 11 participants, comprising five women and six men, with ages ranging from 23 to 31 years old. During the data acquisition process, volunteers are seated in a chair positioned in front of a table where the PC is located. The LMC is mounted above the table, and participants perform the tasks by moving their hands approximately 30 cm above the device. The acquisition set up is shown in Figure 2, also indicating the field of view and the suggested minimum distance from the device. The volunteers underwent a 15 minute acquisition procedure in which they performed the required hand movements, namely OC and TF, for each hand. The acquisition order for each hand is as follows:

- OC Right
- OC Wrong
- TF Right
- TF Wrong

Each execution is recorded for a duration of 20 seconds. Therefore, the handcraft dataset consists of a total of 88 executions, with 44 executions for each hand.

B. PREPROCESSING

Before the data processing, the frame rate during each acquisition is verified. This is done to ensure that a consistent frame rate is maintained across all data. The handcraft dataset has an average frame rate of 115.2 Hz, with a standard deviation of 0.1 Hz.

In order to facilitate the analysis of the motor tasks in this procedure, the frame rate of the data has been reduced from approximately 115 Hz to 30 Hz. This reduction in frame rate does not alter the information regarding the movement significantly, as a 30 Hz sampling rate is sufficient to capture the main hand dynamics.

C. DATA AUGMENTATION

To enhance the dataset, we implement a data augmentation technique based on resampling the signals from 0.4 to 1.4 times the original frequency. This approach is employed to generate a wider range of velocities of the same execution. By resampling the signals at varying frequencies, we could simulate different execution speeds, providing the model with a more robust training set that encompasses a broader spectrum of real-world scenarios.

D. FEATURE EXTRACTION

From the spatial coordinates captured by the LMC in each frame, we derived two sets of features: distances and angles. In the following, the subscript represents the finger: 1 for the thumb, 2 for the index finger, 3 for the middle finger, 4 for the ring finger, and 5 for the pinky finger.

The distance features include the following ones.

1) FINGERTIPS TO PALM CENTER DISTANCE (FPD)

The FPD [27], [29] feature (FPD_i) represents the Euclidean distance between the i-th fingertip coordinate ($\overrightarrow{FT_i}$) and the palm coordinate ($\overrightarrow{CP}$).

$$FPD_i = ||\overrightarrow{FT_i} - \overrightarrow{CP}|| \quad \text{for} \quad i = 1, 2, 3, 4, 5 \tag{1}$$

2) THUMBTIP TO OTHER FINGERTIPS DISTANCE (ToFD)

The ToFD feature ($ToFD_j$) represents the Euclidean distance between the j-th fingertip coordinate ($\overrightarrow{FT_j}$) (excluding the thumb) and the thumbtip coordinate ($\overrightarrow{FT_1}$).

$$ToFD_j = ||\overrightarrow{FT_j} - \overrightarrow{FT_1}|| \quad \text{for} \quad j = 2, 3, 4, 5 \tag{2}$$

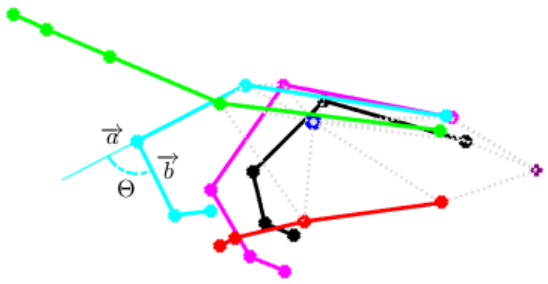

FIGURE 3. Example representation of a joint angle extracted at the joint between proximal phalanges and intermediate phalanges of middle finger.

3) ADJACENT FINGERTIPS DISTANCE (AFD)

The AFD [27], [29], [41] feature (AFD_i) represents the Euclidean distance between the i-th fingertip coordinate ($\overrightarrow{FT_i}$) and its adjacent fingertip coordinate ($\overrightarrow{FT_{i+1}}$).

$$AFD_i = ||\overrightarrow{FT_i} - \overrightarrow{FT_{i+1}}|| \quad \text{for} \quad i = 1, 2, 3, 4 \tag{3}$$

Here, $\overrightarrow{FT_i}$, and $\overrightarrow{CP}$ are three-dimensional coordinates in the LMC coordinate system.

The angle features include the following ones.

4) JOINT ANGLE (JA)

The JA [19], [26], [28], [41] feature represents the angle between two adjacent bones at each finger joint, specifically from the metacarpals to the distal phalanges bones. For example, the JA value of the joint between the proximal phalanges and intermediate phalanges of the middle finger can be calculated as follows:

$$JA = \arccos\left(\frac{\vec{a} \cdot \vec{b}}{|\vec{a}||\vec{b}|}\right) \tag{4}$$

where $\vec{a}$ and $\vec{b}$ are the vectors representing the adjacent bones at a finger joint, as shown in Figure 3. The JA value is computed for each finger joint. To identify the specific joint and finger associated with the computed JA, two indices are added: $JA_{i,j}$, where i represents the finger (from 1 to 5), and j represents the joint (from 1 to 3, except for the thumb, which has 2 joints). This indexing scheme allows for distinguishing between different fingers and their corresponding joints when analyzing the JA values.

5) ADJACENT FINGERTIPS ANGLE (AfA)

The AfA [19], [27], [28], [29], [41] feature represents the angle between every two adjacent fingertip vectors, which corresponds to the angle between the vectors from the palm center to the fingertip coordinates. The computation of AfA values is based on Equation (4) and can be expressed as:

$$AfA_i = \arccos\left(\frac{\overrightarrow{CT_i} \cdot \overrightarrow{CT_{i+1}}}{|\overrightarrow{CT_i}||\overrightarrow{CT_{i+1}}|}\right) \quad \text{for} \quad i = 1, 2, 3, 4 \tag{5}$$

where, $\overrightarrow{CT_i}$ and $\overrightarrow{CT_{i+1}}$ refer to the vectors that represent the distances between the center of the palm and the coordinates of the i-th fingertip and the adjacent fingertip, respectively. These vectors are computed in the LMC coordinate system and have three-dimensional components.

According to the literature, the distances features have been shown to be valuable in capturing the spatial relationships and distances between various hand landmarks, aiding in hand gesture recognition and analysis, while the angles features have proven to be effective in capturing the angular relationships between different finger joints and fingertips, making them suitable for hand gesture recognition tasks.

Hand size variation poses a significant challenge in hand gesture recognition, as highlighted in previous studies [45]. To address this issue and account for the differences in hand sizes among the patients, the features were normalized. Specifically, the distance features were divided by the cumulative Euclidean distance between the palm center and the middle finger tip, denoted as L (also see Figure 4):

$$L = |\overrightarrow{MC}| + |\overrightarrow{PP}| + |\overrightarrow{IP}| + |\overrightarrow{DP}| \tag{6}$$

At each sampling point, L was computed by summing the distance between the Proximal Metacarpal Region (PMR) and the index Metacarpophalangeal Joint (MJ), as well as the lengths of all three bones of the middle finger, as indicated in (6), and shown in Figure 4. Note that, for certain hand poses, specifically when the hand is open, the distance between the thumbtip and the fingertip of the ring finger ($ToFD_4$) or pinky finger ($ToFD_5$) can reach values from 0 up to 1.5. As for the other normalized distance features, instead, they belong to the set [0, 1]. To normalize the angle features (JA and AfA) within the range [0, 1], an empirical approach was followed instead. The maximum values observed for JA and AfA in the entire handcraft dataset were used as normalization angles. Specifically, the normalization angle for JA was set to 150 degrees, and the normalization angle for AfA was set to 120 degrees.

Finally, the following feature sets are defined:

$$\Gamma = \{FPD_1, \ldots, FPD_5\} \tag{7}$$
$$\Delta = \{ToFD_2, \ldots, ToFD_5\} \tag{8}$$
$$H = \{AFD_1, \ldots, AFD_4\} \tag{9}$$
$$\Theta = \{JA_{1,1}, \ldots, JA_{5,3}\} \tag{10}$$
$$\Lambda = \{AfA_1, \ldots, AfA_4\} \tag{11}$$
$$\Pi = \{\Gamma, \Delta, H\} \tag{12}$$
$$\Psi = \{\Theta, \Lambda\} \tag{13}$$
$$\Omega = \{\Pi, \Psi\} \tag{14}$$

Π represents the merged feature vector consisting of the distances feature vectors Γ, Δ, and H. It combines the information from FPD, ToFD, and AFD features. Ψ represents the merged feature vector consisting of the angles feature vectors Θ and Λ; it combines the information from JA and AfA features. Ω represents the merged feature vector consisting of both the distances feature vectors Π and the angles feature vectors Ψ; it combines the information from all the previously described features.

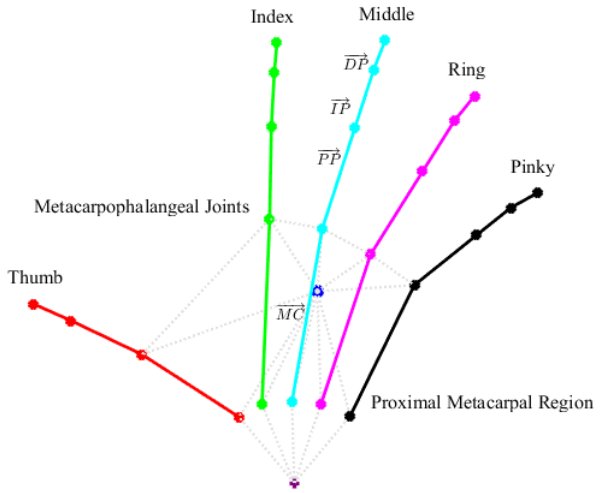

FIGURE 4. Skeletal representation to clarify the anatomy of the PMR and the MJs. The PMR is the connecting area between the metacarpal bones and the carpal bones of the wrist. On the other hand, the MJs are the joints that connect the metacarpal bones to the proximal phalanges of the fingers.

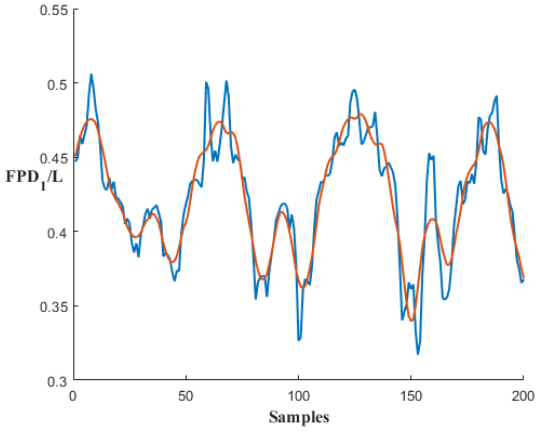

FIGURE 5. Example of signal filtering by Savitzky-Golay filter. The raw thumb FPD values from the execution of a complex motor task are in blue, while the filtered thumb FPD values are shown in orange.

E. FILTERING

To enhance the signal-to-noise ratio while minimizing signal distortion, the Savitzky-Golay filter was employed [46]. Figure 5 displays a 6.6 seconds segment of the FPD feature extracted from the thumb during the execution of the complex motor task when filtered with a Savitzky-Golay filter of second order and frame length of 19 samples. After applying the Savitzky-Golay filter, the signal undergoes significant noise reduction, resulting in a more reliable representation of the thumb's movement during a complex motor task.

F. CLASSIFICATION

The classification of the execution is performed using a RNN, namely, a LSTM in two variants: unidirectional and bidirectional. This model is well-suited for sequence learning tasks, and its use has become increasingly popular in recent works [19], [26], [27], [28], [29], [39], [41]. For online implementation, the output mode of each LSTM layer is always set to "sequence" to achieve sequence-to-sequence classification, meaning that the LSTM provides a classification output for each input sample. The next step is to determine the hyperparameters for training the Uni-LSTM and Bi-LSTM model. Three common parameters are considered: batch size, number of epochs, and number of hidden layer units. The hyperparameters used during the training phase are selected based on the testing accuracy following a coarse grid search, and their best values are provided in Table 1. Additionally, a 50% dropout rate is applied to prevent overfitting.

TABLE 1. Parameter settings.

Hyperparameters	Uni-LSTM	Bi-LSTM
Batch Size	60	60
Number of Epochs	1000	2500
Number of Hidden States	30	30
Dropout Rate	50%	50%

In the MATLAB environment, we have chosen the "classificationlayer" element as the loss function. This function computes the cross-entropy loss for classification tasks with mutually exclusive classes, using the following expression [47]:

$$\text{loss} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{K} w_i t_{ni} \ln y_{ni} \tag{15}$$

where N represents the number of samples in the dataset, K represents the number of classes, w_i is the weight for class i, t_{ni} is an indicator variable that represents whether the n-th sample belongs to the i-th class, and y_{ni} is the output of the model for the n-th sample and the i-th class. Specifically, y_{ni} represents the probability that the network assigns to the n-th input belonging to the i-th class, and it is computed using the softmax function. The softmax function transforms the output of the model into a probability distribution over the classes, ensuring that the probabilities sum up to 1. The loss function, based on the cross-entropy, is a commonly used objective function in classification tasks, aiming to minimize the difference between the predicted probabilities and the true class labels. The Adam optimizer [48] is selected for training the models. The learning rate is fixed at 0.0001 throughout the entire training phase. Training is performed using an NVIDIA RTX 2060 GPU, taking approximately 15 minutes to complete 1000 epochs for Uni-LSTMs and 40 minutes to complete 2500 epochs for Bi-LSTMs. The same training on an Intel Core i9-11900K CPU requires approximately 20% more time.

IV. RESULTS

To validate our models, we employed a 10-fold cross-validation technique. In each fold, the dataset was randomly divided into 9 patients for training and 2 patients for

TABLE 2. 10-Fold identification and assessment results on the test set using Bi-LSTMs.

Features	Accuracy (Mean ± StD)	Precision	Sensitivity	F1 score
Γ	75.1 ± 7.3%	79.5%	78.3%	69.4%
Δ	64.6 ± 6.0%	63.0%	54.5%	48.7%
H	72.1 ± 16.0%	70.1%	61.0%	48.0%
Θ	69.8 ± 12.7%	70.8%	68.4%	61.7%
Λ	67.2 ± 14.6%	67.6%	54.1%	47.5%
Π	81.3 ± 10.5%	81.6%	80.7%	73.8%
Ψ	84.2 ± 12.1%	83.7%	84.1%	81.8%
Ω	80.1 ± 9.9%	80.2%	74.2%	72.2%

TABLE 3. 10-Fold identification and assessment results on the test set using Uni-LSTMs.

Features	Accuracy (Mean ± StD)	Precision	Sensitivity	F1 score
Γ	68.2 ± 8.3%	76.2%	74.2%	69.2%
Δ	66.3 ± 8.3%	63.0%	43.0%	29.1%
H	75.3 ± 8.0%	76.4%	71.2%	68.4%
Θ	67.2 ± 7.2%	76.4%	71.2%	68.4%
Λ	63.0 ± 8.2%	61.7%	62.0%	47.6%
Π	74.2 ± 9.2%	77.5%	75.1%	67.8%
Ψ	72.7 ± 13.5%	72.2%	64.6%	59.6%
Ω	75.0 ± 7.8%	76.7%	77.0%	74.0%

validation and testing (33% and 66% for validation and testing, respectively). We evaluated the viability of our approach to identify task execution from the hands of unseen patients using the LSTM models. The patients for the training and validation sets were selected randomly for each fold, ensuring that the same combination of patients was not repeated. We remark that the data in the test set originate from patients who have never been encountered by the model during testing.

To assess the performance of the models, we computed accuracy, precision, recall, and F1 score using the macroaverage values obtained from the confusion matrices. The use of macroaverage values was necessary due to the small imbalance in our dataset, where there were three instances of wrong task executions, two for the right hand OC acquisition and one for the left hand TF acquisition. The results, in terms of mean accuracy and Standard Deviation (StD), for the proposed Bi-LSTM are shown in Table 2. The corresponding results for the Uni-LSTM are described in Table 3. The best-performing features, based on the highest mean accuracy, were also selected to validate our approach using the LMDHG dataset [49]. This dataset comprises dynamic skeletal hand gestures collected from 21 signers, each performing at least one sign, resulting in an average of 13±1 words per signer.

V. DISCUSSION

The evaluation of our proposed models for classifying the two studied tasks involved performing 10-fold cross-validation to calculate various metrics. To provide a comprehensive analysis of the results, the discussion is divided into three sections.

The first part, Section V-A, focuses on analyzing and interpreting the results obtained for the identification and assessment of execution in both motor tasks. This section includes a detailed examination of the metrics presented in Table 2 and Table 3.

Section V-B compares the results in terms of performance and response time, taking into account practical limitations and implementation aspects arising in the AN context.

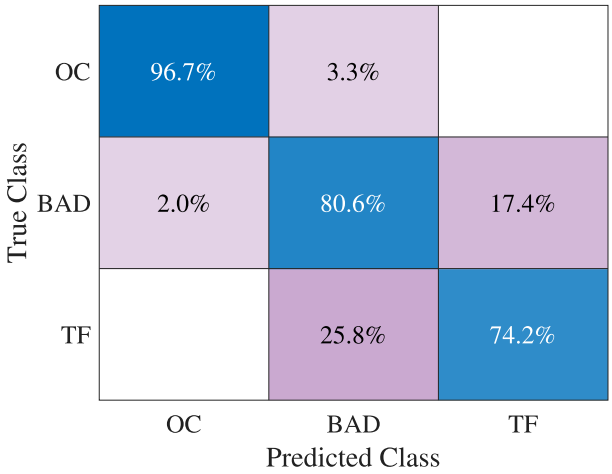

FIGURE 6. Confusion matrix results of Bi-LSTM model for Ψ feature vector.

Lastly, Section V-C presents a comparison with sign language gesture recognition, based on the LMDHG dataset. The results are critically analyzed taking into account feasibility and computational complexity.

A. EXECUTION IDENTIFICATION AND ASSESSMENT

In this section, we report the results obtained using the entire historical signal data, following the 10-fold cross-validation detailed in Section IV. In the following section, instead, we address the problem of real-time application, where only the live data stream is available.

1) Bi-LSTM

Table 2 shows that three feature vectors achieve a mean accuracy equal to or above 80%, namely Π, Ψ, and Ω, with accuracies of 81.3%, 84.2%, and 80.1%, respectively. However, the StD values for almost all feature vectors are greater than 10%, except for Γ (7.3%), Δ (6.0%), and Ω (9.9%).

The cumulative confusion matrix in Figure 6 represents the overall behavior of the Bi-LSTM in the 10-fold cross-validation trials using the Ψ feature vector, which exhibited the highest performance according to the accuracy outlined in Table 2. OC and TF represent correct tasks, while BAD includes OC Wrong and TF Wrong tasks.

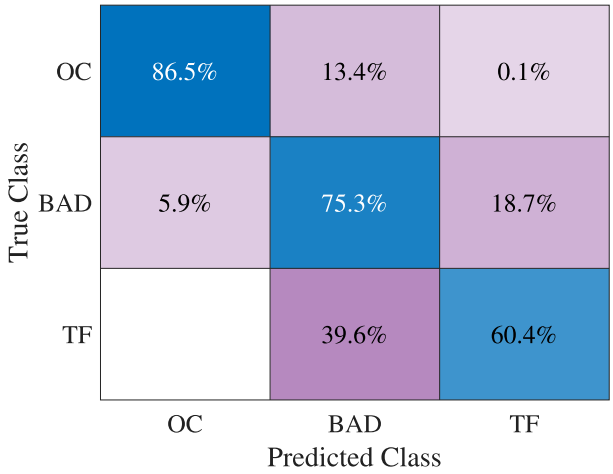

FIGURE 7. Confusion matrix results of Uni-LSTM model for Ω feature vector.

The model demonstrates overall discriminative capability; however, according to the confusion matrix, there is a non-negligible percentage of misclassifications between the TF task and BAD executions. While a correct OC is well classified in 96.7% of the cases, several misclassification errors occur with the TF and the BAD classes. This highlights that the OC task is easier, while accurately identifying and assessing the TF task is challenging. Such a result depends on two factors. First, occlusions in the TF are more critical. In case of a temporary occlusion, two fingers can be swapped in the skeletal model, causing the correct tapping order to appear violated even for a correct TF. This can happen both in training and testing, making classification harder. This issue does not apply to OC, where all the fingers are clenched and spread together. Moreover, the TF execution proves to be more dependent on the person performing the task. Some patients spread the hand wide after each tap, while others tend to keep the hand clenched. Additionally, some patients often tap the index and the pinky twice when inverting the direction. This variability does not occur with the OC, as it is a basic and highly reproducible task.

2) Uni-LSTM

According to the results in Table 3, the AFD feature vector (H), the distance-based (Π), the angle-based (Ψ), and their merging (Ω) achieved a mean accuracy exceeding 70%. Additionally, the StD values for all feature vectors are below 10%, except for Ψ (13.5%). The cumulative confusion matrix in Figure 7 displays the overall performance of the Uni-LSTM across all 10-fold cross-validation trials using the Ω feature vector, which showed the best accuracy as indicated in Table 3.

A higher misclassification rate was observed, up to 39.6%, which is the percentage of TF executions identified as BAD.

As already discussed in the Bi-LSTM case, the OC task is easier to classify compared to the TF task. This is evidenced by the significant difference in correct classification rates: 86.5% for OC versus 60.4% for TF. Such results depend on the same two factor described in the case of Bi-LSTMs.

TABLE 4. Results comparison on LMDHG dataset.

Approach	Accuracy
Lupinetti et al. [50]	91.83%
Bi-LSTM - Ψ	11.2%
Bi-LSTM - Π	8.1%
Bi-LSTM - Ω	9.9%
Uni-LSTM - Ψ	12%
Uni-LSTM - Π	12.6%
Uni-LSTM - Ω	16.6%

B. COMPARISON AND IMPLEMENTATION

Comparing the results of Uni-LSTMs and Bi-LSTMs in the 10-fold cross-validation, the superior accuracy of Bi-LSTMs is evident. However, at the implementation level, Uni-LSTMs and Bi-LSTMs are radically different. Uni-LSTMs can handle data streams because they process data sequentially in one direction, making them suitable for real-time applications. On the contrary, Bi-LSTMs require the entire data sequence to be available to process it in both forward and backward directions. Differently from state-of-the-art applications, such as sign language gesture recognition where a delayed gesture recognition can be accepted, a timely evaluation is crucial in AN. So, in order to apply Bi-LSTMs in real-time, it is necessary to split the data in several chunks to be individually evaluated. Hence, in addition to the computation time, a delay equal to the duration of the chunk must be considered. Moreover, due to the reduced duration of the chunk with respect to the entire movement, a lower accuracy is expected compared to Section V-A1. The size of the chunk length is chosen to comply with a reasonable reaction time during the surgery. Reducing such time interval decreases the evaluation performances, so it represents a trade-off between fast reaction time and accuracy. A two second window has been chosen in the following, representing an acceptable delay and broadly comparable to a neuropsychologist's evaluation and decision time.

Without loss of generality, we employed the features in Ω to verify the feasibility of implementing our approach. The feature vector Ω, which includes both distance-based and angle-based features, was chosen for its lower standard deviation value among the proposed vectors and also because it represents the largest feature vector, thus giving upper bounds on time complexity.

Testing the response time from signal pre-processing to prediction, we found that for every 20 seconds of input signal, the mean time for classifying the entire sequence is 0.5 seconds for Uni-LSTM and 0.6 seconds for Bi-LSTM. With a frame rate of 30 Hz (i.e., 600 samples) in the input sequence, an average computation time of 1 millisecond per

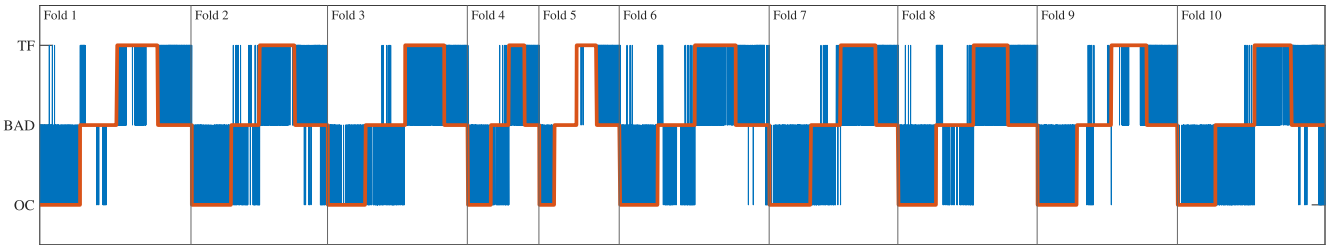

FIGURE 8. Classification results using the feature vector Ω on a data stream with a two second windowing. The orange line represents the true labels, and the blue line represents the Bi-LSTM classification.

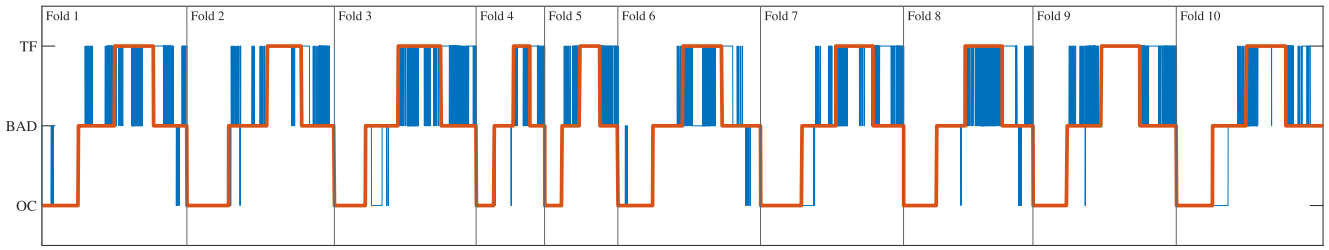

FIGURE 9. Classification results using the feature vector Ω on a data stream. The orange line represents the true labels, and the blue line represents the Uni-LSTM classification.

sample is obtained. Compared to the reaction time of any human neuropsychologist, which is on the order of hundreds of milliseconds, a 1 millisecond delay per sample means that the computation is practically in real-time for the purpose of task identification and assessment. However, the window length delay must be considered for the Bi-LSTMs, in this case 2 seconds, as they need an entire data chunk in input, while Uni-LSTMs do not. Overall, the response time is dominated by the window length, thus making Uni-LSTMs preferable at comparable performance.

Below, we present the testing results obtained from a realistic scenario where only a data stream is available: Uni-LSTMs process the data stream in real-time, while Bi-LSTMs process a delayed acquisition window of two seconds in length. For this analysis, we utilize the best Uni-LSTM and Bi-LSTM networks employing the largest feature vector Ω. The classification results are shown in Figure 8 and Figure 9, where the true labels (OC, TF, and BAD) are represented in orange, and the Uni-LSTM and Bi-LSTM predictions are shown in blue in Figure 8 and Figure 9, respectively. Each fold indicated in Figure 8 and Figure 9 includes the data of the two patients in the test set during the training phase. In the plot of each fold, the data is rearranged by true class for readability.

The relative accuracies using the Bi-LSTM model are 44.5% for OC, 50.3% for TF, and 77.5% for BAD true class. Regarding Uni-LSTM, the relative accuracies for each prediction are 99.% for OC, 89.5% for TF, and 61.9% for BAD. Figure 8 and Figure 9 show the superior performance of the Uni-LSTM when a maximum two seconds delay is enforced on the Bi-LSTM. Please note that, dividing the input data in two second chunks, the accuracy of the best Bi-LSTM is significantly lower than the Bi-LSTM 10-fold average, while the accuracy of the best Uni-LSTM is fully retained and it is better than the Uni-LSTM 10-fold average as expected.

Please note that sporadic misclassification (see for example Fold 2 in Figure 9) could be mitigated with a digital filter, trading off the response time with the accuracy. On the contrary, please note that recurring misclassification occurs especially for TF and BAD. Based on these results, the difficulty in discriminating TF executions from BAD depends heavily on the individual variability in executing the TF task and the criteria used to define a BAD execution.

C. COMPARATIVE ANALYSIS: SIGN LANGUAGE GESTURES

No datasets in the literature deal with hand gesture recognition in AN. To validate the methodology in a different hand gesture recognition task, we test it on a public dataset, namely, LMDHG [49]. The public dataset LMDHG presents a significantly different challenge with respect to the OC and TF identification and assessment, consisting in the classification of 14 different sign language gestures. Table 4 presents the accuracy results on the LMDGH dataset obtained using Uni-LSTM and Bi-LSTM with the most discriminative feature vectors (Π, Ψ, and Ω), compared to a state-of-the-art method [50]. Only the highly discriminant feature vectors Π, Ψ, and Ω, are applied in this comparison for brevity. Despite Π, Ψ, and Ω achieve an accuracy of 81.3%, 84.2%, and 80.1% in OC and TF identification and assessment, when applied to LMDHG hand gesture recognition, Ψ, Π, and Ω obtains only 11.17%, 8.15%, and 9.9%, respectively. Comparable results hold for Uni-LSTMs. In contrast, Lupinetti et al. achieved 91.83% accuracy in the LMDHG dataset. This substantial difference in accuracy can be attributed to the much larger amount of information fed to the CNNs by Lupinetti et al., as well as to the

size of the neural network. Lupinetti et al. [50] employ 1920×1080 resolution images in input, a pre-trained version of a ResNet-50 architecture, trained using a progressive resizing technique and a Ranger neural network optimizer. Please note that the ResNet-50 contains 50 layers and approximately 25.6 million parameters. On the contrary, the proposed method considers only a limited number of features in inputs instead of 1920×1080 images, and the RNNs under investigation count up to 11.9 thousand parameters using a Bi-LSTM and the best feature Ψ. This results in a notable reduction in parameter count, with the proposed LSTM models having a parameter ratio of approximately 1:2000 compared to ResNet-50 for Bi-LSTMs and 1:4000 for Uni-LSTMs. Hence, the proposed network proves to be insufficiently large for the more complex task required by the LMDHG dataset, despite the above 80% accuracy for Bi-LSTMs in OC and TF identification and assessment. On the other hand, the proposed method has lower space and time complexity and proves to be effective for task assessment in AN.

Comparing Table 2 and Table 4, OC and TF identification and assessment proves to be simpler than recognizing among 14 different gesture classes in LMDHG. Also, the timings of the gestures and the pauses in sign language gesture recognition introduce a further increase in difficulty that is not of interest in AN. The results in Table 4 suggest that the accuracy achieved in OC and TF identification, as reported in Table 2, could be further improved with the inclusion of more features and the application of a more complex neural network. However, space and time complexity should be taken into account to decide if any possible increase in accuracy motivates for the use of a more complex method.

VI. CONCLUSION AND FUTURE WORKS

This work introduces a novel approach to support decision-making for neuropsychologists during AN. The study focuses on the classification of two commonly performed hand motion tasks during AN.

The classification is performed using single layer Uni-LSTMs and Bi-LSTMs.

The results indicate that the proposed methodology is capable of discriminating the OC task with a significant true positive rate up to 96.7% in a 10-fold cross-validation, as shown in Figure 6.

However, the classification capability of our method is reduced when it comes to distinguishing TF from BAD executions. The misclassification highlights that, for some patients, the BAD execution data are not sufficiently distinct from TF for a proper classification using the proposed feature vectors. This discrepancy depends on two main factors: camera occlusions and individual differences in task execution. To address camera occlusions, potential solutions include improving the positioning of the camera or employing an array of cameras. Overcoming the individual differences in task execution may involve refining the LSTMs training by including data from the same patient, at least in the case of correct executions.

In the related literature regarding sign language gesture recognition, Bi-LSTMs is preferred over Uni-LSTMs. Accordingly, Bi-LSTMs outperform Uni-LSTMs in AN, using a 10-fold cross-validation, when the entire data sequence is available. However, if we split the data in smaller chunks to meet a maximum response time of two seconds, the accuracy of the best Bi-LSTM drops significantly (44.5% for OC, 50.3% for TF, and 77.5% for BAD), and the best Uni-LSTM is more accurate (99.% for OC, 89.5% for TF, and 61.9% for BAD) than the best Bi-LSTM. When considering the shorter response time of Uni-LSTMs (approximately one millisecond) and the absence of delay due to windowing, this argument renders Uni-LSTMs preferable in the context of AN.

In conclusion, there are three directions for future work. Firstly, developing a larger dataset could help to better capture the nuances of incorrect executions. This would enhance the classification accuracy of BAD executions and reduce the misclassification of BAD executions as TF shown in Figure 6. Secondly, we suggest applying the proposed method in an intraoperative setting to evaluate whether the results during AN in real scenarios are consistent with those obtained from the handcrafted dataset. The expected technical issues in the surgical room consist in positioning the LMC properly. On the contrary, lighting conditions are expected to be marginal because the LMC is equipped with infrared LEDs to illuminate its working area. Such results can then be compared with those obtained using more invasive methods, such as ECoG, considering the possibility of developing a synergistic solution based on both real-time ECoG and LMC to improve classification performance. Thirdly, the opportunity of auto-generating features (hence replacing the ones in Section III-D) directly from images by means of CNNs (e.g., as done in [51]) could be investigated. We remark that the results obtained in the test set regard patients who have never been encountered during testing. This poses several challenges as regards individual differences, such as individual hand size, execution speed, and quality of the execution. The proposed solutions to mitigate these challenges are, respectively, feature normalization, data augmentation, and filtering.

ACKNOWLEDGMENT

The authors would like to thank IEEE Access reviewers for their valuable feedback and insightful comments, which have greatly contributed to the improvement of this manuscript.

REFERENCES

[1] G. Kobyakov, A. Lubnin, A. Kulikov, A. Gavrilov, S. Goryaynov, A. Poddubskiy, and K. Lodygina, "Awake craniotomy," in *Problems of Neurosurgery Named After NN Burdenko*, vol. 1. Moscow, Russia: Media Sphera Publishing House, 2016, pp. 88–96.

[2] S. Bonifazi, C. Passamonti, S. Vecchioni, R. Trignani, P. P. Martorano, V. Durazzi, S. Lattanzi, F. Mancini, and R. A. Ricciuti, "Cognitive and linguistic outcomes after awake craniotomy in patients with high-grade gliomas," *Clin. Neurol. Neurosurgery*, vol. 198, Nov. 2020, Art. no. 106089.

[3] S. Hegde, V. Hande, H. Gunasekaran, A. Shashidhar, and A. Arimappamagan, "Role of clinical neuropsychologists in awake-craniotomy," *Neurol. India*, vol. 69, no. 3, p. 711, 2021.
[4] A. Mansouri, S. Ibrahim, L. Bello, J. Martino, and C. Velasquez, "The current state of the art of primary motor mapping for tumor resection: A focused survey," *Clin. Neurol. Neurosurgery*, vol. 229, Jun. 2023, Art. no. 107685.
[5] S. L. Hervey-Jumper and M. S. Berger, "Introduction: Surgical management of eloquent area tumors," *Neurosurgery*, vol. 87, no. 6, pp. 1076–1077, 2020.
[6] E. Kahn, M. Lane, and O. Sagher, "Eloquent: History of a word's adoption into the neurosurgical lexicon," *J. Neurosurgery*, vol. 127, no. 6, pp. 1461–1466, Dec. 2017.
[7] G. Lepski, J. Honegger, M. Liebsch, M. G. Sória, P. Narischat, K. F. Ramina, T. Nägele, U. Ernemann, and M. Tatagiba, "Safe resection of arteriovenous malformations in eloquent motor areas aided by functional imaging and intraoperative monitoring," *Operative Neurosurgery*, vol. 70, pp. ons276–ons289, Jun. 2012.
[8] A. T. Wang, P. Pillai, E. Guran, H. Carter, T. Minasian, J. Lenart, and R. Vandse, "Anesthetic management of awake craniotomy for resection of the language and motor cortex vascular malformations," *World Neurosurgery*, vol. 143, pp. e136–e148, Nov. 2020.
[9] G. Brydges, R. Atkinson, M. J. Perry, D. Hurst, T. Laqua, and J. Weimers, "Awake craniotomy: A practice overview," *AANA J.*, vol. 80, no. 1, pp. 61–68, Feb. 2012.
[10] R. H. Muster, J. S. Young, P. Y. M. Woo, R. A. Morshed, G. Warrier, S. Kakaizada, A. M. Molinaro, M. S. Berger, and S. L. Hervey-Jumper, "The relationship between stimulation current and functional site localization during brain mapping," *Neurosurgery*, vol. 88, no. 6, pp. 1043–1050, 2021.
[11] M. A. Morrison, F. Tam, M. M. Garavaglia, L. Golestanirad, G. M. T. Hare, M. D. Cusimano, T. A. Schweizer, S. Das, and S. J. Graham, "A novel tablet computer platform for advanced language mapping during awake craniotomy procedures," *J. Neurosurgery*, vol. 124, no. 4, pp. 938–944, Apr. 2016.
[12] N. U. F. Hameed, Z. Zhao, J. Zhang, L. Bu, Y. Zhou, L. Jin, H. Bai, W. Li, J. Tang, J. Lu, J. Wu, and Y. Mao, "A novel intraoperative brain mapping integrated task-presentation platform," *Operative Neurosurgery*, vol. 20, no. 5, pp. 477–483, Apr. 2021.
[13] N. Sanai, Z. Mirzadeh, and M. S. Berger, "Functional outcome after language mapping for glioma resection," *New England J. Med.*, vol. 358, no. 1, pp. 18–27, Jan. 2008.
[14] B. Navarro-Main, L. Jiménez-Roldán, P. González Leon, A. M. Castaño-León, A. Lagares, and Á. Pérez-Nuñez, "Neuropsychological management of awake patient surgery: A protocol based on 3 years' experience with glial tumours," *Neurocirugía English Ed.*, vol. 31, no. 6, pp. 279–288, Nov. 2020.
[15] F. Weichert, D. Bachmann, B. Rudak, and D. Fisseler, "Analysis of the accuracy and robustness of the leap motion controller," *Sensors*, vol. 13, no. 5, pp. 6380–6393, May 2013.
[16] P. E. Kim and M. Singh, "Functional magnetic resonance imaging for brain mapping in neurosurgery," *Neurosurgical Focus*, vol. 15, no. 1, pp. 1–7, Jul. 2003.
[17] T. Xie, D. Zhang, Z. Wu, L. Chen, and X. Zhu, "Classifying multiple types of hand motions using electrocorticography during intraoperative awake craniotomy and seizure monitoring processes—Case studies," *Frontiers Neurosci.*, vol. 9, pp. 1–10, Oct. 2015.
[18] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[19] D. Avola, M. Bernardi, L. Cinque, G. L. Foresti, and C. Massaroni, "Exploiting recurrent neural networks and leap motion controller for the recognition of sign language and semaphoric hand gestures," *IEEE Trans. Multimedia*, vol. 21, no. 1, pp. 234–245, Jan. 2019.
[20] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.
[21] M. Oelschlägel, T. Meyer, U. Morgenstern, H. Wahl, J. Gerber, G. Reiß, E. Koch, G. Steiner, M. Kirsch, G. Schackert, and S. B. Sobottka, "Mapping of language and motor function during awake neurosurgery with intraoperative optical imaging," *Neurosurgical Focus*, vol. 48, no. 2, pp. 1–10, p. E3, Feb. 2020.
[22] K. Volkova, M. A. Lebedev, A. Kaplan, and A. Ossadtchi, "Decoding movement from electrocorticographic activity: A review," *Frontiers Neuroinform.*, vol. 13, p. 74, Dec. 2019.
[23] M. Khademi, H. M. Hondori, A. McKenzie, L. Dodakian, C. V. Lopes, and S. C. Cramer, "Free-hand interaction with leap motion controller for stroke rehabilitation," in *Proc. CHI Extended Abstr. Human Factors Comput. Syst.*, Apr. 2014, pp. 1663–1668.
[24] A. Kaur and S. Bansal, "Deep learning for dynamic hand gesture recognition: Applications, challenges and future scope," in *Proc. 5th Int. Conf. Multimedia, Signal Process. Commun. Technol. (IMPACT)*, Nov. 2022, pp. 1–6.
[25] M. Saad, T. Yang, and H. Zhou, "A comparison of bidirectional GRU and LSTM for hand gesture recognition using leap motion," in *Proc. 37th Youth Academic Annu. Conf. Chin. Assoc. Autom. (YAC)*, Nov. 2022, pp. 1427–1433.
[26] A. Alnahhas, B. Alkhatib, N. Al-Boukaee, N. Alhakim, O. Alzabibi, and N. Ajalyakeen, "Enhancing the recognition of Arabic sign language by using deep learning and leap motion controller," *Int. J. Sci. Technol. Res.*, vol. 9, no. 4, pp. 1865–1870, 2020.
[27] L. Yang, J. Chen, and W. Zhu, "Dynamic hand gesture recognition based on a leap motion controller and two-layer bidirectional recurrent neural network," *Sensors*, vol. 20, no. 7, p. 2106, Apr. 2020.
[28] S. B. Abdullahi and K. Chamnongthai, "American sign language words recognition using spatio-temporal prosodic and angle features: A sequential learning approach," *IEEE Access*, vol. 10, pp. 15911–15923, 2022.
[29] C. K. M. Lee, K. K. H. Ng, C.-H. Chen, H. C. W. Lau, S. Y. Chung, and T. Tsoi, "American sign language recognition and training method with recurrent neural network," *Expert Syst. Appl.*, vol. 167, Apr. 2021, Art. no. 114403.
[30] J. Wu, N. Yu, Y. Yu, H. Li, F. Wu, Y. Yang, J. Lin, J. Han, and S. Liang, "Intraoperative quantitative measurements for bradykinesia evaluation during deep brain stimulation surgery using leap motion controller: A pilot study," *Parkinson's Disease*, vol. 2021, pp. 1–7, Jun. 2021.
[31] A. H. Butt, E. Rovini, C. Dolciotti, P. Bongioanni, G. De Petris, and F. Cavallo, "Leap motion evaluation for assessment of upper limb motor skills in Parkinson's disease," in *Proc. Int. Conf. Rehabil. Robot. (ICORR)*, Jul. 2017, pp. 116–121.
[32] A. H. Butt, E. Rovini, C. Dolciotti, G. De Petris, P. Bongioanni, M. C. Carboncini, and F. Cavallo, "Objective and automatic classification of Parkinson disease with leap motion controller," *Biomed. Eng. OnLine*, vol. 17, no. 1, pp. 1–21, Nov. 2018.
[33] M. J. Kim, E. Naydanova, B. Y. Hwang, K. A. Mills, W. S. Anderson, and Y. Salimpour, "Quantification of Parkinson's disease motor symptoms: A wireless motion sensing approach," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 3658–3661.
[34] R. Rastgoo, K. Kiani, and S. Escalera, "Hand sign language recognition using multi-view hand skeleton," *Expert Syst. Appl.*, vol. 150, Jul. 2020, Art. no. 113336.
[35] S. Sharma and K. Kumar, "ASL-3DCNN: American sign language recognition technique using 3-D convolutional neural networks," *Multimedia Tools Appl.*, vol. 80, no. 17, pp. 26319–26331, Jul. 2021.
[36] S. B. Abdullahi, K. Chamnongthai, V. Bolon-Canedo, and B. Cancela, "Spatial–temporal feature-based end-to-end Fourier network for 3D sign language recognition," *Expert Syst. Appl.*, vol. 248, Aug. 2024, Art. no. 123258.
[37] H. B. D. Nguyen and H. N. Do, "Deep learning for American sign language fingerspelling recognition system," in *Proc. 26th Int. Conf. Telecommun. (ICT)*, Apr. 2019, pp. 314–318.
[38] S. B. Abdullahi and K. Chamnongthai, "IDF-sign: Addressing inconsistent depth features for dynamic sign word recognition," *IEEE Access*, vol. 11, pp. 88511–88526, 2023.
[39] S. B. Abdullahi and K. Chamnongthai, "American sign language words recognition of skeletal videos using processed video driven multi-stacked deep LSTM," *Sensors*, vol. 22, no. 4, p. 1406, Feb. 2022.
[40] D. Naglot and M. Kulkarni, "Real time sign language recognition using the leap motion controller," in *Proc. Int. Conf. Inventive Comput. Technol. (ICICT)*, vol. 3, Aug. 2016, pp. 1–5.
[41] H. Sharif, A. Eslaminia, P. Chembrammel, and T. Kesavadas, "Classification of activities of daily living based on grasp dynamics obtained from a leap motion controller," *Sensors*, vol. 22, no. 21, p. 8273, Oct. 2022.
[42] P. Chophuk, K. Pattanaworapan, and K. Chamnongthai, "Fist American sign language recognition using leap motion sensor," in *Proc. Int. Workshop Adv. Image Technol. (IWAIT)*, Jan. 2018, pp. 1–4.
[43] T.-W. Chong and B.-G. Lee, "American sign language recognition using leap motion controller with machine learning approach," *Sensors*, vol. 18, no. 10, p. 3554, Oct. 2018.

[44] P. Chophuk and K. Chamnongthai, "Backhand-view-based continuous-signed-Letter recognition using a rewound video sequence and the previous signed-letter information," *IEEE Access*, vol. 9, pp. 40187–40197, 2021.
[45] N. M. Bhiri, S. Ameur, I. Alouani, M. A. Mahjoub, and A. B. Khalifa, "Hand gesture recognition with focus on leap motion: An overview, real world challenges and future directions," *Expert Syst. Appl.*, vol. 226, Sep. 2023, Art. no. 120125.
[46] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, Jul. 1964.
[47] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*, vol. 4. New York, NY, USA: Springer, 2006.
[48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[49] S. Y. Boulahia, E. Anquetil, F. Multon, and R. Kulpa, "Dynamic hand gesture recognition based on 3D pattern assembled trajectories," in *Proc. 7th Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Nov. 2017, pp. 1–6.
[50] K. Lupinetti, A. Ranieri, F. Giannini, and M. Monti, "3D dynamic hand gestures recognition using the leap motion sensor and convolutional neural networks," in *Proc. Int. Conf. Augmented Reality, Virtual Reality Comput. Graph.* Cham, Switzerland: Springer, Mar. 2020, pp. 420–439.
[51] K. Lai and S. N. Yanushkevich, "CNN+RNN depth and skeleton based dynamic hand gesture recognition," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 3451–3456.

LUIGI GABRIEL TROCONIS (Student Member, IEEE) was born in Caracas, Venezuela, in 1998. He received the B.S. and M.S. degrees in biomedical engineering from the Università Politecnica delle Marche, Ancona, Italy, in 2020 and 2022, respectively, where he is currently pursuing the Ph.D. degree in information engineering. He is also a member of the Laboratory of Artificially Intelligent Robotics, Università Politecnica delle Marche. His research interests include computer vision, machine learning, pattern recognition, and robotics.

RICCARDO FELICETTI (Member, IEEE) received the master's degree (cum laude) in computer and automation engineering and the Ph.D. degree (cum laude) in information engineering from the Università Politecnica delle Marche, in 2016 and 2021, respectively. Since 2019, he has been a Postdoctoral Research Fellow with the Università Politecnica delle Marche. His main research interests include fault detection and diagnosis, fault tolerant control, and optimization with applications to unmanned vehicles and energy management systems.

ANDREA MONTERIÙ (Member, IEEE) received the M.Sc. degree in electronic engineering and the Ph.D. degree in artificial intelligence systems from the Università Politecnica delle Marche, Italy, in 2003 and 2006, respectively. Currently, he is an Associate Professor with the Department of Information Engineering, Università Politecnica delle Marche, where he leads the Laboratory of Artificially Intelligent Robotics (former Robotics Laboratory). He has edited 11 books and he has published more than 200 scientific papers in the area of fault diagnosis, fault-tolerant control, robotics, periodic systems, control of intelligent systems, assistive robotics, and human–machine interaction. He has participated in more than 20 research projects funded by public and private entities. His research interests include fault diagnosis, fault tolerant control, nonlinear dynamics and control, periodic and stochastic system control, and applied in different fields, including aerospace, marine, robotic and unmanned artificial intelligent systems, and assistive systems for healthcare and wellbeing. He has served as the Vice-Chair for the IEEE Italy Section CE Soc Chapter and as the Chair for the IEEE CTSoc Technical Committee on Consumer Systems for Healthcare and Wellbeing.

Open Access funding provided by 'Università Politecnica delle Marche' within the CRUI CARE Agreement