Seizure: European Journal of Epilepsy 65 (2019) 65–71

Contents lists available at ScienceDirect

Seizure: European Journal of Epilepsy

journal homepage: www.elsevier.com/locate/seizure

Aberrant epileptic seizure identification: A computer vision perspective

David Ahmedt-Aristizabal[a,*], Clinton Fookes[a], Simon Denman[a], Kien Nguyen[a], Sridha Sridharan[a], Sasha Dionisio[b]

[a] Image and Video Research Laboratory, SAIVT, Queensland University of Technology, Australia
[b] Department of Mater Advanced Epilepsy Unit, Mater Centre for Neurosciences, Brisbane, Australia

ARTICLE INFO

Keywords:
Semiology
Aberrant behavior
Seizure motion libraries
Computer vision
Deep learning

ABSTRACT

Purpose: The recent explosion of artificial intelligence techniques in video analytics has highlighted the clinical relevance in capturing and quantifying semiology during epileptic seizures; however, we lack an automated anomaly identification system for aberrant behaviors. In this paper, we describe a novel system that is trained with known clinical manifestations from patients with mesial temporal and extra-temporal lobe epilepsy and presents aberrant semiology to physicians.

Methods: We propose a simple end-to-end-architecture based on convolutional and recurrent neural networks to extract spatiotemporal representations and to create motion capture libraries from 119 seizures of 28 patients. The cosine similarity distance between a test representation and the libraries from five aberrant seizures separate to the main dataset is subsequently used to identify test seizures with unusual patterns that do not conform to known behavior.

Results: Cross-validation evaluations are performed to validate the quantification of motion features and to demonstrate the robustness of the motion capture libraries for identifying epilepsy types. The system to identify unusual epileptic seizures successfully detects out of the five seizures categorized as aberrant cases.

Conclusions: The proposed approach is capable of modeling clinical manifestations of known behaviors in natural clinical settings, and effectively identify aberrant seizures using a simple strategy based on motion capture libraries of spatiotemporal representations and similarities between hidden states. Detecting anomalies is essential to alert clinicians to the occurrence of unusual events, and we show how this can be achieved using pre-learned database of semiology stored in health records.

1. Introduction

The analysis of clinical signs such as facial modifications (*e.g.* blinking, chewing, smacking), limb automatisms, ictal head turning and hand movements (*e.g.* hand dystonia, tapping, grabbing) [1,2], may provide clues as to the cerebral networks underpinning the epilepsy. However, the study of these signs relies heavily on clinical experience and training. Given the importance of body motion patterns in the assessment of epilepsy, prior works have demonstrated that automated analysis of semiological patterns based on computer vision can support diagnosis by standard and objective assessment methods among evaluators [3,4]. Previously, we proposed a facial semiology analysis approach to identify patients with mesial temporal lobe epilepsy (MTLE) [5], and a hierarchical multi-modal system to quantify and classify patients with MTLE and extra-temporal lobe epilepsy (ETLE) based on face, body and hand motions [6].

One limitation of these approaches is their reliance on supervised learning, which assumes the test data originates from one of the training categories. When unseen data (*i.e.* data for a new patient) does not come from one of the training categories, the trained model is of no use. This is evident in low performance when classifying the semiology of unseen patients with potential aberrant or unusual seizures (leave-one-subject-out cross-validation scheme) [6]. In this scenario, the ictal patterns from the test patients with MTLE or ETLE are not strongly correlated to the semiological patterns of other patients previously observed and contained within the training and validation data. Therefore, the system will show low classification accuracies for new patients with aberrant or unusual semiological features, even if these patients suffer from the same condition.

Aberrant identification methods can be very useful in identifying interesting, concerning, or unknown events using past patient cases stored in health records. The identification of anomalies corresponding to unusual semiology will alert clinicians to consider different diagnostic choices. Once a deviation is detected, it may be used to generate

* Corresponding author.
E-mail address: david.aristizabal@hdr.qut.edu.au (D. Ahmedt-Aristizabal).

https://doi.org/10.1016/j.seizure.2018.12.017
Received 6 November 2018; Received in revised form 14 December 2018; Accepted 18 December 2018

a patient-specific alert for consideration by clinicians.

Our strategy for aberrant identification is by grouping patients into a best-fit model, using a template from a pre-learned database of known semiology in the form of libraries. The libraries store feature representations of the motion exhibited by the patient during the recorded seizure. These motion libraries enable us to identify if the semiology from a new patient can fit into the status quo of the learned information, to conclude on having dissimilar findings or aberrant semiology.

In order to conduct this process, we train a deep learning system that detects, tracks, and captures clinical manifestations of known behaviors for two types of seizure (MTLE and ETLE). These motions, which are represented by spatiotemporal features (characteristics of shape and motion in videos) and extracted from the trained system, are saved in motion capture (MoCap) libraries. We adopt these libraries to identify if new data samples belong to either of these known semiologies, or represent anomalous behaviors. Our approach is based on two key intuitions: (1) Deep learning and computer vision have revolutionized human motion understanding, producing accurate motion features of particular dynamics, and (2) MoCap libraries are useful for distinguishing behaviors through simple memorization of previously observed behavior and similarities between features. The technical contributions of our work are summarized as follows:

1) We propose a novel and simple end-to-end architecture based on Convolutional Neural Networks (CNN) and a Long Short-Term memory (LSTM) architecture, to extract deep spatiotemporal features that are used to construct the MoCap libraries of known semiologies with limited training data.
2) We present the first method in the literature that identifies aberrant semiology during epileptic seizures by comparing with features of known behaviors recorded in the MoCap libraries using the simple and effective discriminative cosine similarity measure.

The remainder of this paper is organized as follows: Section 2 discusses our proposed approach and presents the dataset. In each of the subsections, the intuition and reasoning behind our approach are explained. Section 3 explains the experimental approach and results, and Section 4 discusses the results. Finally, Section 5 concludes the paper.

2. Methods

In this paper, we introduce a system that is able to determine whether a test patient has unusual semiological patterns that do not conform to known behaviors stored in health records. We design an architecture that quantifies semiology to develop MoCap libraries which are used to identify aberrant epileptic seizures. To achieve this, we propose a system with the structure as presented in Fig. 1.

Each video clip captures one seizure from a patient with MTLE or ETLE, to develop two MoCap libraries (MTLE and ETLE) of known behaviors. To quantify semiology, the inputs of the system are short video sequences rather than a whole video, such that we obtain more data to train the system. We define a sequence as 25 consecutive frames. We preprocessed this dataset by detecting the patient and resizing the images in all seizures. We train a CNN–LSTM structure [7] in a supervised fashion to model the relationship between known semiologies with different epilepsy types. Then, spatiotemporal representations from all sequences are extracted from an LSTM layer [8] to generate MoCap libraries for each type of epilepsy. Lastly, when evaluating a new seizure (test patient), we split the recorded seizure into smaller segments (sequences) and match each to the library. We adopt the cosine similarity metric to compute the similarity of each sequence from the test patient with all the sequences of each library. As a result of this, we consider a threshold of acceptance based on the total number of sequences that are similar to the library to identify aberrant behavior. A cross-validation evaluation is performed to verify the flexibility of the system to capture and quantify human motion behavior, and to demonstrate the sensitivity and specificity of each MoCap library to distinguish known behaviors. We attempt to discriminate between clinical presentations, which may help to address limitations of current automated approaches for semiology assessment [6] by incorporating the identification of unusual clinical manifestations within the system. Details of each phase are described in the following subsections.

2.1. Data collection and specifications

Video recordings during EEG and Stereo-EEG were captured as a part of the routine long-term monitoring protocol with patients with drug-resistant epilepsy at the Mater Hospital in Brisbane, Australia; a tertiary referral public epilepsy surgery center. These patients were monitored over a time period ranging from 5 to 7 days (epilepsy surgery cases with high seizure frequency at baseline). Each video recording represents the seizure event via the captured semiology. Each seizure video was segmented such that it showed from the first epileptic discharge until the full expression of semiology prior to version and convulsion, if it was experienced. This results in videos clips of roughly 1–2 min length. We include as many natural clinical settings as possible to ensure a challenging database that reflects real conditions, which includes existing collections of retrospective clinical data and patients that were under evaluation during, approximately, the past two years.

The two categories of MTLE and ETLE, were chosen in order to try and categorize the complexity of semiology into two defined regions. The term “Extra-temporal” was based on Stereo-EEG localization of seizures not arising from the mesial temporal structures. Thus seizures arising from the insula or opercular regions were deemed extra-temporal in origin, even if they eventually involved the temporal lobe later as part of the ictal network. While MTLE can also present with complicated semiology, it generally has been shown to have a more limited repertoire of signs and thus is more easily distinguished from seizures arising from other regions [9]. These more homogeneous features assist with the machine learning in computer vision and allow easier comparison.

A total of 119 seizures from 14 patients (62 seizures) with MTLE and 14 patients (57 seizures) with ETLE were used to create the motion capture libraries of known behaviors, where the number of seizures per patient are depicted in Tables 2 and 3 . This data represents a total of 102,225 frames or 4089 sequences (25 consecutive video frames per sequence). In the MTLE group 12 patients underwent a Stereo-EEG as they were lesion negative on MRI. All 14 patients in the ETLE group underwent Stereo EEG. Seizure freedom was seen in all 28 cases to add further weight to localization.

To evaluate the aberrant seizure identification approach, 5 patients separate from the group used to create the MoCap libraries of known behaviors were selected, all of whom were deemed to have unusual semiology. These seizures were selected as their clinical manifestations were not similar to other cases, or show deviations to baselines that are well described and are reproducible in the majority of diagnosed patients, *i.e.* they were different to the baseline of what most patients experience. According to Table 4, the aberrant semiology for these patients is described as follow: Patient 1 and 2 exhibit fear expressions, Patient 3 presents swallowing motions, Patient 4 demonstrates finger snapping, and Patient 5 turns their body along the horizontal axis.

2.2. Patient detection

To extract features related to the patients’ behavior, we first define the region of interest that contains the patient. This ensures that the majority of features used in the kinematic analysis come from the patient and not from any family members or physicians also visible in the videos. This procedure also helps deal with different camera-bed viewing angles and changes in the inclination angle of the bed. This

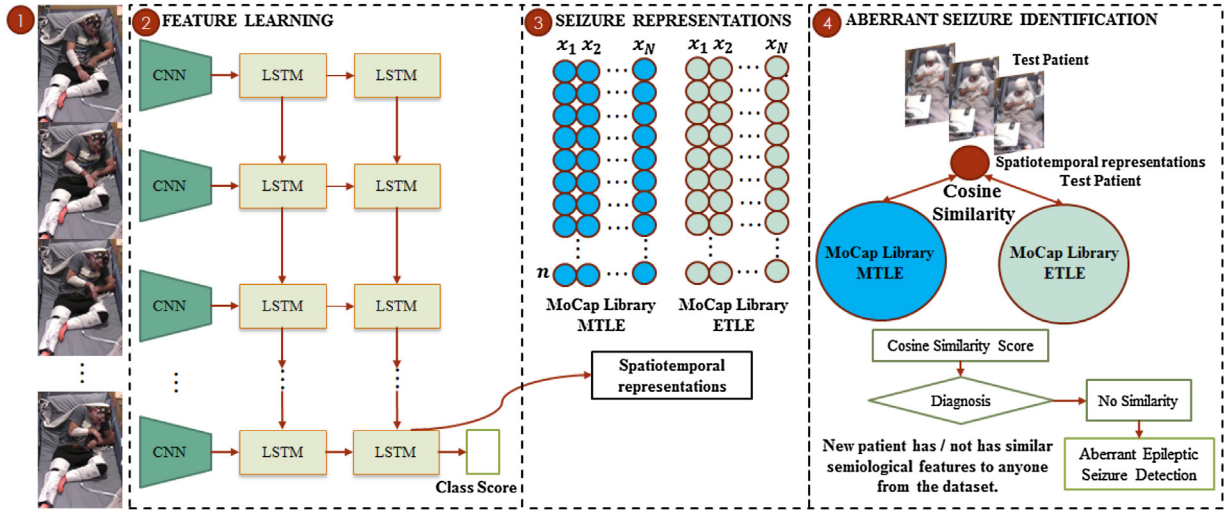

Fig. 1. Overview of the proposed framework used to identify aberrant semiology based on MoCap libraries. (1) Each video clip captures one seizure from a patient. The inputs of the system are sequences defined as 25 consecutive frames. Detection of the patient is performed to improve the accuracy in quantifying clinical manifestations. (2) CNN: Convolution Neural Network; LSTM: Long-Short Term Memory; A CNN-LSTM architecture is designed and trained to quantify and distinguish known behaviors from two types of epilepsy (MTLE and ETLE). (3) Spatiotemporal representations are extracted and temporally-constructed into feature matrices that represent a MoCap library of each type of seizure; $x_1 \cdots x_N$ corresponds to each seizure in each known group with N total seizures, and n indicates the number of sequences. (4) The cosine similarity measure is used to identify the similarity between a test patient and each MoCap library. If a new patient has dissimilar semiology from patients in the dataset, the patient is considered an aberrant epileptic case.

region is defined as the location of the bed that contains the patient, and it is expected that the patient will remain inside this boundary during a seizure. Detection is performed only in the first frame, to ensure that the features for each video segment are extracted from the same region of the frame.

Among object detection processes, state-of-the-art methods are driven by deep neural networks. We perform object boundary detection using the Mask-RCNN architecture [10], which is a benchmark object detection method trained on the COCO dataset [11]. In our approach, we use pretrained Mask-RCNN weights to perform human and bed detection. We expand the detected patient bounding box with an offset of 20% of the total width on each side to avoid the extremities of the patient being located outside of this boundary due to movements during a seizure. Then, we crop and resize all images of each sequence to a resolution of 550 × 720 pixels. A selected sequence of images with the patient detected is illustrated in Fig. 1.

2.3. Deep learning architecture and training

The network architecture use to create the MoCap libraries is based upon well-known cascaded networks for action recognition [7], and works by capturing spatiotemporal features from video sequences to predict classes through an end-to-end deep learning model. Cascaded networks are proposed to first extract discriminative representations from static images (CNNs), and then input these features to sequential networks (LSTMs). Although current computer vision approaches are moving into deeper networks, we design a shallow hybrid network using the CNN–LSTM architecture due to the limited data available. Shallow CNN architectures have also already been shown to be suitable for epileptic seizure detection with a small amount of training data [12]. We exploited many insights about suitable network architectures. It could be argued that using benchmark architectures such as AlexNet and VGG to extract spatial information may be preferable over the proposed shallow CNN architecture. However, we choose to design and train our own network to enable our model to learn representations of discriminative patterns of seizure disorders using a small dataset. It is highly recommended to train small networks (150K parameters in our design) because we found experimentally that state-of-the-art networks are not suitable due to their large number of training parameters (AlexNet-60M; VGG-144M) and the multiple classes they seek to categorize (the ImageNet database has 1000 object categories). In our scenario, using networks with such a large number of parameters quickly led to over-fitting, or to no significant improvement, even when using regularization techniques such as overlap pooling, image augmentation and dropout [13].

Through extensive experiments, we explore different design choices for our model. The design of the network architecture that shows the best performance for the task of learning spatiotemporal features is displayed in Fig. 2.

For training and testing, all images extracted in the patient detection phase are resized to 155 × 200 pixels. A feature map is extracted from each input sequence by a CNN architecture containing: (1) one convolutional layer with stride 1 and 8 units, (2) one LeakyReLU activation layer, (3) one normalization layer, (4) one max pooling layer with stride 12 and (5) one fully connected layer with ReLU activation and 64 units. The CNN output is subsequently fed to a stacked LSTM architecture. We adopt an LSTM with 2 hidden layers of 128 and 64 units respectively [5]. Finally, the output of the second hidden recurrent layer is fed into a densely connected layer with a sigmoid activation function to describe the probability of each sequence having MTLE or ETLE behavior.

We train the CNN–LSTM network by optimizing the binary cross-entropy loss using the Adam optimizer [14] with a learning factor of 10^{-3}, and the first and second moments decay rates of 0.9 and 0.999, respectively. We adopt a batch-size of 32 and train the model over 30 epochs using the default initialization parameters from Keras [15] for initializing the weights of the hidden units. We use the Theano backend [16], and balance the training data at the sequence level using the class weight parameters.

2.4. Motion capture libraries (MoCap libraries)

The dynamics or representation of semiology to develop the MoCap within each sequence are extracted from the deep learning architecture via the LSTM layer with 64 hidden units as shown in Figs. 1 and 2 . Once the features for each sequence are extracted, the representation for each library has a dimensionality of [2697, 64] and [1392, 64] of known semiologies from patient with MTLE and ETLE, respectively. The

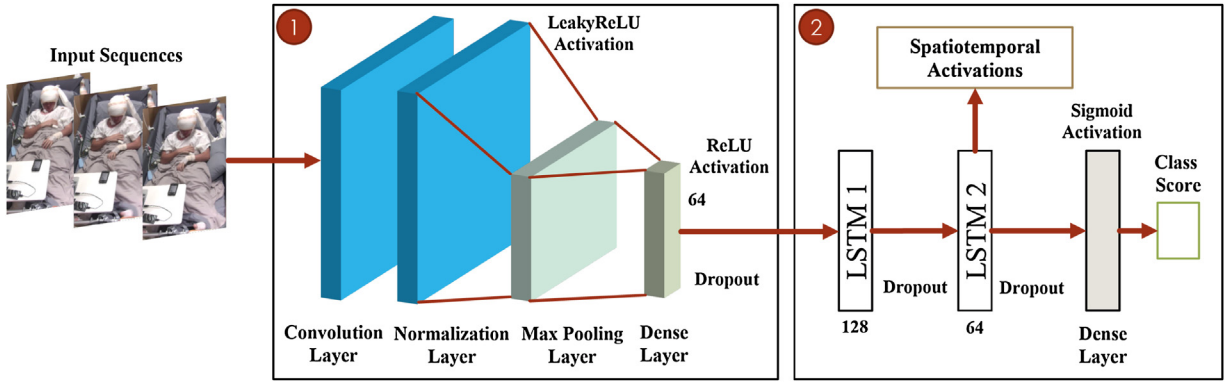

Fig. 2. The end-to-end network architecture to identify epileptic patients and extract spatiotemporal representations. (1) CNN architecture: one convolutional layer with eight [5 × 5] filter kernels and LeakyReLU activations; one normalization layer; a max pooling layer with pooling size [12 × 12]; a flatten and dense layer with 64 output units with ReLU activation; dropout of 50%. (2) LSTM architecture: 2 hidden layers of 128 and 64 hidden units respectively with dropout of 35% after each LSTM layer to reduce overfitting. We perform classification using a densely connected layer with a sigmoid activation function. Temporal activations are extracted from the LSTM layer to develop MoCap libraries of semiological features.

first dimension indicates the number of sequences for all seizures, and the second dimension refers to the number of spatiotemporal features extracted from the CNN–LSTM architecture. When evaluating a new seizure, we split the test seizure into sequences and match each sequence to the library. The trained CNN–LSTM architecture is used to extract a feature vector from each sequence, and this feature vector is used to compute similarities to the libraries.

2.5. Identifying aberrant epileptic behavior

The cosine similarity is used to measure how alike two data samples are, *i.e.* if the new seizure has/does not have similar features to any sequence of semiology included in each MoCap library. The cosine similarity is a widely used metric for measuring the similarity between hidden states and it is more effective for discriminating the hidden states of deep neural networks than traditional methods using Jacquard similarity or SVMs [17]. The similarity measure is a distance with dimensions representing features of the objects. If this distance is small, it indicates a high degree of similarity while a large distance indicates a low degree of similarity. The similarity is defined as,

$$\text{Similarity} = 1 - \cos(\theta), \tag{1}$$

Given two vectors of attributes $A = [x_1, x_2, x_n]$ and $B = [y_1, y_2, y_n]$, the $cos(\theta)$ is the measure of the angle between the two vectors and is given by:

$$\cos(\theta) = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}, \tag{2}$$

Given m spatiotemporal feature sequences in one MoCap library and t_p spatiotemporal features sequences from a test patient, the result of the cosine similarity will have a representation of size $[m, t_p]$. The values of the cosine similarity range between -1 and 1, where 1 indicates vectors in the same direction ("similar") and -1 indicates vectors in the opposite directions ("dissimilar"). We calculate the average similarity for each sequence of the test patient with each MoCap library in order to identify aberrant behavior. The average similarity for t_p sequences of a test patient can be represented as $[1, t_p]$. Given this average similarity per sequence, we adopt two experimental boundaries: the first one determines if a single sequence is similar to the MoCap library. The second one, determines if a test patient has aberrant clinical manifestations based on the total number of sequences from the patient's entire that are similar to each MoCap library. To decide that a single sequence of a test patient is similar to an entry in the MoCap library, we require a cosine similarity average score to exceed a threshold of 0.6. Then, we calculate the total number of sequences that are similar to the library, to determine a percentage of acceptance. If the percentage of acceptance, *i.e.* the total number of sequences that are similar to the MoCap library, is less than 30%, we consider the test patient to have aberrant or unusual epileptic seizures. This means that the features of the test patient do not conform to well-known behavior. These boundary levels were defined based on visual inspection of the results, where the main aim was to identify when the similarity was unclear. If during the process of aberrant identification there is evidence of similarity between sequences of the test patient and a MoCap library, the system can indicates the patient that belong those sequences included in the dataset that shows a degree of *similarity* > 0.6 from strong to weak similarity.

3. Results

3.1. Experimental setup

We adopt a k-fold cross-validation in order to validate the flexibility of the system to capture, quantify and model the variations in the data compared to different approaches that capture human motion behaviors. In this scenario, the sequences used for validation and testing are completely separate to those used for training the model, but it is possible to have sequences from the same patient in each set. The sequences of all patients of the same class are split into 70% for training, 15% for validation, 15% for testing and k is set to 10 [18]. The alternative approaches are based on a two-stage architecture. The first stage extracts spatial features using well-known methods which extract human motion features, Mask-RCNN and optical flow, in the same images used in the training and test of our shallow model (155 × 200 pixels). In the second stage, the spatial features are fed to our LSTM architecture which exploits the dynamic variation of these features. Using the architecture of Mask-RCNN [10] and the segmentation of the detected patient, spatial features are extracted from the last fully connected layer in order to capture the semiology. These features have a dimension of [1, 4096], as the output of the layer in the network has 4096 units. Each sequence of features of the movements has a dimensionality of [25, 4096], capturing 25 frames, each with 4096 features. In the case of features from optical flow, we identify the motion vectors related to the human body motions. We compute the optical flow between adjacent frames using FlowNetv2 [19]. We use one threshold on the flow to ensure that there is motion in the frame, *i.e.* more than 10% pixels have optical flow values above zero. The performance of established networks such as VGG16 and ResNet coupled with our LSTM design shows AUC values under 0.5, which indicates that the models were not training. For this reason, these architectures

are not considered further to compare the performance of the proposed system.

We adopt a leave-one-subject-out cross-validation (LOSO-CV) scheme in order to demonstrate the sensitivity and specificity of each MoCap library to distinguish seizures from the two different groups of known epilepsy. For each patient diagnosed with the same type of epilepsy as the MoCap library, we compared the acceptance of the patient with the CNN-LSTM architecture prediction. The LOSO-CV approximates a real clinical scenario when analyzing the entire video corpus for a test patient, who is totally excluded from the training data. This results in an unbiased estimation of the true generalization error.

In order to test the capability of each MoCap library to identify aberrant seizures based on the cosine similarity, we selected 5 patients in this experiment that exhibited seizures with unusual clinical manifestations as discussed in Section 2.1.

3.2. Verification of MoCap libraries and identification of aberrant seizures

The CNN–LSTM model achieved an average AUC of 0.9703 for the *k*-fold cross-validation scheme. This result outperforms alternative strategies based on baseline approaches which extract human motion features. The performance of these models are summarized in Table 1.

Tables 2 and 3 display the identification performance conducted for each patient group using each MoCap library, and compares the LOSO-CV performance of the CNN-LSTM architecture for patients of the same group. The “Proportion sequences” column indicates the proportion of the entire corpus that is from each subject. Verification performance using the trained architecture is shown in the “Test Accuracy CNN-LSTM” column, which reached an average accuracy of 66.48% and 62.19% for patients with MTLE and ETLE, respectively. These results represent a large variability in the verification of each patient because of the high variation in the data when classifying the semiology of particular patients, which is one of the significant limitations of such strategies [6].

On the other hand, the identification of seizures based on the proposed MoCap libraries and average similarities show more consistent results, as can be seen by the acceptance test of each patient. We have included the average of the cosine similarity of all sequences for each test patient. A label of “Yes” indicates that the patient is considered similar to the library because the percentage of sequences from all seizures with a cosine similarity > 0.6 is greater than 30%. Similarly, a label of “No” indicates that the patient is unrelated to the dataset. The acceptance level of the MTLE MoCap library indicates that 13 of 14 patients were correctly categorized as MTLE and 12 of 14 patients with ETLE were unrelated with the library (Sensitivity: 92.85%, Specificity: 85.71%). Similarly, 11 of 14 patients with ETLE and 13 of 14 patients with MTLE were properly identified with the ETLE MoCap (Sensitivity: 78.57%, Specificity: 92.85%). These results confirm that the MoCap libraries are more robust for identifying clinical presentations than a trained classifier approach, and can be used for the experimental phase of identifying unusual seizures.

The performance when analyzing test patients with aberrant semiologic features is displayed in Table 4. The system successfully identifies 4 of the 5 seizures categorized as aberrant semiology correctly. This result allows us to demonstrate that a simple strategy based on MoCap libraries and similarities between features is a promising technique to identify clinical manifestations that can be unusual compared to well-described semiology. From the five selected patients in this experiment, patient 3 was considered similar to the MTLE library, which indicates a false positive detection. Although for this test patient the average similarity of all sequences was low (0.2017), the total number of sequences that were similar to the MoCap library was 34.5%, which is higher than the defined threshold of 30%. For this reason, the test patient was not considered as an aberrant epileptic seizure. In this situation, we have indicated the patients from the MoCap library that were found to have similar motions to patient 3 based on the level of similarity. It is possible to argue that the swallowing motion of patient 3 has a similar dynamic to the subtle mouth motions of patients M11 and M12.

Table 1
Multifold cross-validation performance with alternative approaches.

Approach	Validation accuracy (%)	Test accuracy (%)	AUC
Mask-RCNN + LSTM	72%	60%	0.68
Opticalflow + LSTM	90%	75%	0.903
CNN–LSTM	**93.4%**	**90%**	**0.9703**

4. Discussion

We have developed an automated system that is able to identify aberrant seizures utilizing visual cues from known behaviors (*e.g.* facial expression, limb posturing, repetitive movements, etc). When the semiology of a new patient does not fit the status-quo of learned information, would be considered as having dissimilar findings, thus aberrant semiology.

Semiology evaluation is dependent on observer experience and training. Such experience is developed over a long period of time by clinicians and through the comparison of past individual patients. In this contribution, the knowledge learned through previous patients is generalized and described in terms of MoCap libraries of known behaviors, that were extracted from a trained shallow deep learning architecture.

From the experimental evaluations conducted, we have shown that the presented simple end-to-end deep architecture for semiology quantification, *i.e.*, detection and tracking of body movement patterns, and MoCap library construction, is robust to the challenging imaging conditions typical of an EMU. By using a shallow network, the system can be more flexible and simple to transfer to low-end devices such as portable or embedded devices. Our study shows that the analysis of patients considering all body motions simultaneously is viable using the existing monitoring technology in the hospital. As it was confirmed in previous studies [5,6], the performance of LOSO-CV schemes is strictly related to semiological patterns contained in the dataset. This is evident from the difference in classification performance between patients based on the deep learning architecture. The obtained results showed that the identification performance ranges from 36.25% to 100% for MTLE and from 30.52% to 100% for ETLE. On the other hand, the MoCap library performance has revealed a consistent performance by categorizing the majority of the patients correctly with an average sensitivity of 85.71% and specificity of 89.28% between the two libraries. By demonstrating that is possible to detect outlier semiologies, it is possible to perform active learning by including these new features in a system trained to classify epileptic seizures [6] to avoid incorrect interpretation during diagnosis.

We have demonstrated the benefits of our CNN-LSTM architecture and argue that it has advantages over baseline approaches which extract human motion features such as Mask-RCNN [10] and optical flow [19]. Approaches that have proposed the use of optical flow estimation with deep networks for action recognition have shown encouraging results. Although this technique can be used to measure seizures that involve limb and head movements, we argue that they are unsuitable for detecting subtle movements related to semiology from facial modifications and hand motions. Additionally, the strategy of using the segmentation mask [10] of the detected patient can be unsuitable because this segmentation cannot capture the totality of the patient's body with a fine-tuned model.

We argue that the identification of anomalies is essential to alert clinicians to the occurrence of unusual events that deviate from the majority of examples recorded in the hospital. However, this work is far from being able to replace the expertise of clinical practise. It is an

Table 2
Identification performance based on the MTLE MoCap library. We compare MTLE patients data to the MTLE MoCap Library and compute their similarity to an MTLE model trained using the CNN–LSTM method (Section 2.3). For completeness, we also compare each of the ETLE sequences to the MTLE MoCap library.

Patient MTLE	Number seizures	*f* (Proportion sequences (%)	Test accuracy CNN–LSTM (%)	Average cosine similarity	Test acceptance MoCap library	Patient ETLE	Average cosine similarity	Test acceptance MoCap Library
M1	4	6.12	66.40	0.8390	Yes	E1	−0.1017	No
M2	3	5.47	78.63	0.8382	Yes	E2	−0.3180	No
M3	3	2.38	100	0.8390	Yes	E3	−0.2490	No
M4	8	15.90	48.24	0.5549	Yes	E4	0.0150	No
M5	5	11.08	43.46	0.5420	Yes	E5	0.1050	No
M6	3	4.96	53.77	0.3490	Yes	E6	−0.2540	No
M7	6	10.38	50.00	0.3349	Yes	E7	0.3290	Yes
M8	2	1.03	63.64	0.6248	Yes	E8	−0.0100	No
M9	7	12.39	77.36	0.7125	Yes	E9	0.1930	No
M10	1	2.34	100	0.8390	Yes	E10	0.2700	No
M11	5	7.48	36.25	0.1117	No	E11	0.1580	No
M12	10	13.93	46.64	0.4620	Yes	E12	−0.2850	No
M13	4	4.63	85.86	0.7120	Yes	E13	−0.3115	No
M14	1	1.92	80.49	0.7380	Yes	E14	0.2865	Yes
Average			**66.48**		**13/14**			**12/14**

attempt to use a novel approach based on computer vision to take on a complex area such as seizure semiology. The proposed techniques as it currently stands, is not for precise localization purposes or to define the exact underlying epileptic network, but rather to detect semiologies which may be considered “outliers”, thereby triggering the need for further investigations. For example, in the case of an MRI lesion and epilepsy, the lesion itself may be unreliable and semiology would be the key to proceed with the diagnosis.

We would also argue that the proposed approach is flexible enough to support finer granularity in term of diagnostic assistance as more data becomes available. Our methodology can be extended to create motion libraries with more specific localizations. With our current system, lateralizing signs (head version, figure 4, contralateral dystonia, post ictal nose wipe, etc.) were not seen in all of the seizures and lateralizing features varied between patients making it difficult to isolate these specific features. At present we have insufficient data to investigate this, however we have shown that the approach to identify aberrant seizures is promising, suggesting that with more data the system can achieve greater utility.

This work is by no means a complete solution to semiology, but rather a completely novel method, unreported in the literature, to tackle this highly complex area through further research. We expect that our results will provide the basis of technological development for encoding semiology in the form of MoCap libraries. This would enable more efficient transfer learning between clinical experts and hospitals, avoiding the ethical problem of using identifiable information of patients. When a new patient is presented, his/her seizure can be easily searched for a match in the database conditioned on the behavior in the MoCap library. It may also help to encourage the adoption of new treatment targets based on recommendations directly from the similarity between patients that were successfully diagnosed.

5. Conclusions

This paper has proposed a novel approach to identify aberrant epileptic behaviors based on a simple strategy using motion capture (MoCap) libraries extracted from an end-to-end deep learning architecture, and similarities to pre-learned semiology. The proposed computer vision based technique which attempts new clinical representations would be valuable to clinicians in the pre-surgery assessments of epileptic patients by alerting them to the occurrence of unusual events that deviate from the majority of examples recorded in the hospital. We envision these methods will enable knowledge discovery, where unusual outcomes and their contexts are identified. This research effectively exploits existing camera monitoring infrastructure to modernize Epilepsy Monitoring Units without incurring additional costs, by automating many manual monitoring and analysis tasks.

Table 3
Identification performance based on the ETLE MoCap library. We compare ETLE patients data to the ETLE MoCap Library and compute their similarity to an ETLE model trained using the CNN-LSTM method (Section 2.3). For completeness, we also compare each of the MTLE sequences to the ETLE MoCap library.

Patient ETLE	**Number Seizures**	**Proportion sequences (%)**	**Test Accuracy CNN-LSTM (%)**	**Average Cosine Similarity**	**Test Acceptance MoCap Library**	**Patient MTLE**	**Average Cosine Similarity**	**Test Acceptance MoCap Library**
E1	3	4.48	96.67	0.8059	Yes	M1	−2.0050	No
E2	7	7.46	100	0.8390	Yes	M2	−1.9000	No
E3	3	5.22	100	0.8390	Yes	M3	−1.7580	No
E4	1	2.76	67.30	0.5248	Yes	M4	−0.1060	No
E5	7	12.31	40	0.4915	Yes	M5	0.0180	No
E6	2	1.57	100	0.8390	Yes	M6	0.1090	No
E7	3	7.31	45.71	0.2415	Yes	M7	0.2120	No
E8	4	11.94	63.75	0.8215	Yes	M8	−0.1080	No
E9	6	17.09	40.17	0.4798	Yes	M9	−0.0045	No
E10	5	4.93	36.36	0.1011	No	M10	−0.1790	No
E11	6	11.49	30.52	0.0569	No	M11	0.1060	Yes
E12	4	6.27	57.36	0.8390	Yes	M12	0.0050	No
E13	3	4.18	55.36	0.8390	Yes	M13	−2.1060	No
E14	3	2.99	37.5	0.0958	No	M14	−1.8090	No
Average			62.19		11/14			13/14

Table 4
Aberrant epileptic seizure identification based on each MoCap library.

Test	Number	Label	MTLE MoCap library		ETLE Mocap library		Possible similar
Patient	**Seizures**		**Average Cosine Similarity**	**Is an aberrant patient?**	**Average Cosine Similarity**	**Is an aberrant patient?**	**Patients**
1	1	Aberrant	0.1117	**Yes**	−0.0190	**Yes**	
2	1	Aberrant	0.0056	**Yes**	−0.1240	**Yes**	
3	1	Aberrant	**0.2017**	No	−0.1390	**Yes**	**M11, M12**
4	1	Aberrant	0.0124	**Yes**	0.1240	**Yes**	
5	1	Aberrant	0.1070	**Yes**	−0.2540	**Yes**	

Conflict of interest statement

The authors report no conflicts of interest.

Acknowledgments

This research was partly supported by Mater Centre for Neuroscience under Research Governance authorization RG-17-008 and the Australian Research Council Discovery Grant DP140100793. We confirm that we have read the journal's position on issues involved in ethical publication, and this report is consistent with those guidelines. The experimental procedures involving human subjects described in this paper were approved by the Mater Health Services Human Research Ethics Committee.

References

[1] Noachtar S, Peters AS. Semiology of epileptic seizures: a critical review. Epilepsy BehaV 2009;15(1):2–9.
[2] Chauvel P, McGonigal A. Emergence of semiology in epileptic seizures. Epilepsy BehaV 2014;38:94–103.
[3] Pediaditis M, Tsiknakis M, Leitgeb N. Vision-based motion detection, analysis and recognition of epileptic seizures – a systematic review. Comput Methods Programs Biomed 2012;108(3):1133–48.
[4] Ahmedt-Aristizabal D, Fookes C, Dionisio S, Nguyen K, Cunha JPS, Sridharan S. Automated analysis of seizure semiology and brain electrical activity in presurgery evaluation of epilepsy: a focused survey. Epilepsia 2017;58(11):1817–31.
[5] Ahmedt-Aristizabal D, Fookes C, Nguyen K, Denman S, Sridharan S, Dionisio S. Deep facial analysis: a new phase I epilepsy evaluation using computer vision. Epilepsy BehaV 2018;82:17–24.
[6] Ahmedt-Aristizabal D, Fookes C, Denman S, Nguyen K, Sridharan S, Dionisio S. A hierarchical multi-modal system for motion analysis in epileptic patients. Epilepsy BehaV 2018;87:46–58.
[7] Donahue J, Anne Hendricks L, Guadarrama S, Rohrbach M, Venugopalan S, Saenko K, Darrell T. Long-term recurrent convolutional networks for visual recognition and description. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2015:2625–34.
[8] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735–80.
[9] Maillard L, Vignal J-P, Gavaret M, Guye M, Biraben A, McGonigal A, Chauvel P, Bartolomei F. Semiologic and electrophysiologic correlations in temporal lobe seizure subtypes. Epilepsia 2004;45(12):1590–9.
[10] He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. Proceedings of the IEEE International Conference on Computer Vision (ICCV) 2017:2980–8.
[11] Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. Microsoft COCO: common objects in context. Proceedings of the European Conference on Computer Vision (ECCV). 2014. p. 740–55.
[12] Achilles F, Tombari F, Belagiannis V, Loesch AM, Noachtar S, Navab N. Convolutional neural networks for real-time epileptic seizure detection. Comput Methods Biomech Biomed Eng: Imaging Vis 2016:1–6.
[13] Pasupa K, Sunhem W. A comparison between shallow and deep architecture classifiers on small dataset. 2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE). 2016. p. 1–6.
[14] Kingma DP, Adam JBa. A Method for Stochastic Optimization. 2014. arXiv preprint arXiv:1412.6980.
[15] Chollet F. Keras. 2015.
[16] Al-Rfou R, Alain G, Almahairi A, Angermueller C, Bahdanau D, Ballas N, Bastien F, Bayer J, Belikov A, Belopolsky A, et al. Theano: A Python Framework for Fast Computation of Mathematical Expressions. 2016. arXiv preprint arXiv:1605.02688.
[17] Fernando T, Denman S, Sridharan S, Fookes C. Tracking by prediction: a deep generative model for mutli-person localisation and tracking. Proceeding of the IEEE Winter Conference on Applications of Computer Vision (WACV). 2018. p. 1122–32.
[18] Kohavi R, et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. Proc Intl Joint Conf Artif Intell 1995;14(2):1137–45.
[19] Ilg E, Mayer N, Saikia T, Keuper M, Dosovitskiy A, Brox T. Flownet 2.0: evolution of optical flow estimation with deep networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2 2017.