Moro Matteo (Orcid ID: 0000-0002-8644-0175)
Nobile Giulia (Orcid ID: 0000-0002-9957-7760)
Gibbs Steve Alex (Orcid ID: 0000-0002-8489-4907)
Nobili Lino (Orcid ID: 0000-0001-9317-5405)

Automatic Video Analysis and Classification of Sleep-related Hypermotor seizures and Disorders of Arousal

Matteo Moro[1,2,3,*], Vito Paolo Pastore[1,2,*], Giorgia Marchesi[1,3,4], Paola Proserpio[5], Laura Tassi[5], Anna Castelnovo[6,7,8], Mauro Manconi[6], Giulia Nobile[9], Ramona Cordani[9,10], Steve A. Gibbs[11], Francesca Odone[1,2,3], Maura Casadio[1,3], Lino Nobili[3,9,10**]**

[1] Department of Informatics, Bioengineering, Robotics and Systems Engineering (DIBRIS), University of Genova, 16146, Genova, Italy
[2] Machine Learning Genoa (MaLGa) Center, University of Genova, 16146, Genova, Italy
[3] RAISE Ecosystem, Genova, Italy
[4] Movendo Technology, 16128, Genova, Italy
[5] "C. Munari" Epilepsy Surgery Centre, Niguarda Hospital, Milan, Italy
[6] Sleep Medicine Unit, Neurocenter of Southern Switzerland, Ospedale Civico, Lugano, Switzerland
[7] Faculty of Biomedical Sciences, Università Della Svizzera Italiana, Lugano, Switzerland
[8] University Hospital of Psychiatry and Psychotherapy, University of Bern, Bern, Switzerland
[9] Child Neuropsychiatry Unit, IRCCS Istituto Giannina Gaslini, Member of the ERN EpiCARE, 16147, Genova, Italy
[10] Dipartimento di Neuroscienze, Riabilitazione, Oftalmologia, Genetica e Scienze Materno-Infantili (DINOGMI), University of Genova, 16146, Genova, Italy
[11] Department of Neurosciences, Center for Advanced Research in Sleep Medicine, Sacred Heart Hospital, University of Montreal, Montreal, Quebec, Canada.

* Equally first contributing authors
** Equally senior contributing authors

Abstract

Objective. Sleep-related hypermotor epilepsy (SHE) is a focal epilepsy with seizures occurring mostly during sleep. SHE seizures present different motor characteristics ranging from dystonic posturing to hyperkinetic motor patterns, sometimes associated with affective symptoms and complex behaviors. Disorders of Arousal (DOA) are sleep disorders with paroxysmal episodes that may present analogies with SHE seizures. Accurate interpretation of the different SHE patterns and their differentiation from DOA manifestations can be difficult, expensive, and require highly skilled personnel not always available. Furthermore, it is operator dependent.

Methods. Common techniques for human motion analysis, such as wearable sensors (*e.g.*, accelerometers) and motion capture systems, have been considered to overcome these problems. Unfortunately, these systems are cumbersome and they require trained personnel for markers and sensors positioning, limiting their use in the epilepsy domain. To overcome these problems, recently, a lot of effort has been spent in studying automatic methods based on video analysis for the characterization of human motion. Systems based on computer vision and deep learning have been exploited in many fields, but epilepsy has received limited attention.

Results. In this paper we present a pipeline composed by a set of 3D Convolutional Neural Networks that, starting from video recordings, reached an overall accuracy of 80% in the classification of different SHE semiology patterns and DOA.

Significance. The preliminary results obtained in this study highlighted the fact that our deep learning pipeline could be used by physicians as a tool to support them in the differential diagnosis of the different patterns of SHE and DOA, and encourage for further investigation.

Keywords: Epilepsy Detection, Video analysis, Deep Learning, Disorders of Arousal, Sleep Hypermotor Epilepsy

Correspondence

Matteo Moro, Department of Informatics, Bioengineering, Robotics and Systems Engineering (DIBRIS) and Machine Learning Genoa (MaLGa) Center, University of Genova, 16146, Genova, Italy. Email: matteo.moro@edu.unige.it

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/epi.17605

Highlights

- Sleep-related hypermotor epilepsy (SHE) is a focal epilepsy with seizures occurring during sleep and with various motor characteristics
- Disorders of Arousal (DOA) are sleep disorders with paroxysmal episodes that may present analogies with SHE seizures
- The semiologic interpretation of SHE episodes and their differential diagnosis with DOA require costly resources and specific expertise
- We implemented a machine learning pipeline based on video analysis able to discriminate among different SHE semiology patterns and DOA
- Our pipeline successfully discriminated different SHE semiology patterns and DOA with a test accuracy of 80%

Funding information

This work was funded by the European Union - NextGenerationEU. However, the views and opinions expressed are those of the authors alone and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

This work was supported by MNESYS (PE0000006) – A Multiscale integrated approach to the study of the nervous system in health and disease (DN. 1553 11.10.2022)

This work was supported by the Italian Ministry of Health, 5 x 1000 project 2017, "SOLAR: Sleep disorders in children: an innovative clinical research perspective".

V.P.P. was supported by FSE REACT-EU-PON 2014–2020, DM 1062/2021.

Conflict of interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses and interpretation of data; in the writing of the manuscript and in the decision to publish the results.

Ethical publication

We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

Data availability statement

The data presented in this study are available only on request from the corresponding author. The data are not publicly available due to privacy reasons.

Ethics approval statement

The current retrospective study received the approval of the Niguarda Hospital ethics committee (ID 939–12.12.2013).

Patient consent statement

All patients gave written consent for the use of video recordings for research and publications purposes.

1 | INTRODUCTION

Sleep-related hypermotor epilepsy (SHE) is a focal epilepsy with seizures occurring mostly during sleep. Seizure semiology in people with SHE shows wide heterogeneity in terms of duration and complexity, depending on the brain regions involved. Indeed, ictal manifestations can consist in brief or sustained dystonic posturing but also in hyperkinetic motor patterns, sometimes associated with affective symptoms or ambulatory behaviors [1]. Seizures in SHE may have a frontal or an extra-frontal origin [2, 3, 4], and an accurate localization of seizure origin is of crucial importance for candidates to epilepsy surgery. Seizures in SHE have been classified into four clusters of semiology patterns (SP) namely: SP1, elementary motor signs; SP2, unnatural hypermotor movements; SP3, integrated hypermotor movements and SP4, gestural behaviors with high emotional content [4]. When dealing with people with a possible SHE, the challenge may lie not only in understanding where seizures originate from, but also in differentiating these manifestations from other sleep-related paroxysmal episodes, such as Disorders of Arousal (DOA). DOA are characterized by involuntary motor manifestations of various complexities, where bizarre and violent behaviors may occur during an incomplete awakening from sleep [5]. The differential diagnosis between SHE and DOA may be achieved through anamnestic collection, the use of questionnaires and the analysis of paroxysmal episodes recorded at home (home video-recordings) or in the laboratory (Video-Polysomnography) [6, 7, 8, 9, 10]. However, both the semeiologic interpretation of SHE episodes and their differential diagnosis with DOA are time consuming and may require costly examinations, resources and specific expertise. For these reasons, in the last few years, the promising results reached with automated and semi-automated movement analysis techniques opened up a new frontier in the field of SHE.

One of the first attempts in the development of automatic systems to analyze and detect epileptic seizures (ES) was performed with the use of accelerometers [11]. In this direction, many studies adopted accelerometers to automatically detect ES or psychogenic nonepileptic seizures (PNES) [12, 13, 14]. The results obtained with the use of these sensors have been proven to be effective, but, unfortunately, they may suffer from bias due to ambiguity on the position of the accelerometers on the body [12]. Furthermore, the presence of wearable sensors during sleep may result in high discomfort levels. For these reasons, other techniques have been considered in the detection, analysis and classification of motor manifestations of various complexities [15]. A state-of-the-art methodology to analyze motion patterns largely used in the medical domain relies on motion capture systems and markers [16]. Chen et al. [17] used the trajectories over time of the markers placed all over participants bodies to classify between frontal (hyperkinetic, tonic posturing, fencing posture, tonic head turning), temporal lobe and psychogenic nonepileptic seizures. However, these techniques also present many disadvantages [18, 19]. In fact, markers are intrusive, they may limit natural movements and their location must be assigned *a priori* by expert operators, making the overall procedure time-consuming and operator-dependent [18].

For these reasons, video analysis has been recently considered as a possible alternative to marker-based systems and wearable sensors to characterize human motion [20, 21, 22] and extract quantitative information in an automatic and less invasive modality. This option was made possible by the increasing progress - in terms of accuracy and performance – of deep learning algorithms in addressing video analysis [23]. Nowadays video analysis has been exploited to quantitatively analyze human motion [24, 25] with promising results in many fields, including the medical domain (*e.g.*, physical rehabilitation [26, 27]). However, few studies until now have been performed in the epilepsy field [28, 29, 30].

Recently, Karacsony et al. [31] combined infrared and depth videos (acquired with a stereo system) to implement a multimodal approach to differentiate between epileptic seizures in frontal lobe epilepsy, temporal lobe epilepsy and non-epileptic events.

Ahmedt-Aristizabal et al. [29] proposed an approach based on two different levels. First, by using pose estimation algorithms [25] (*i.e.*, computer vision methods that include the detection of semantic keypoints and the definition of the body skeleton), they detected the positions of keypoints on the body of the person in the scene. Starting from this information, they computed quantitative kinematic parameters (landmark-based approach) and they used them to classify seizures. In this first approach, they also computed optical flow [32, 33] (*i.e.*, a computer vision method used to detect and to characterize the direction and the intensity of moving parts in images) to enrich the information obtained from the pose. Secondly, they developed what they called a region-based approach, an end-to-end pipeline that leveraged deep learning algorithms (*e.g.*, 2D Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) [34]) to characterize different motion patterns.

Other examples that leveraged video-based and machine learning-based systems to detect and characterize major nocturnal motor seizures were presented by Peltola et al. [35] and Larsen et al. [36]. In these works, the authors presented systems to automatically detect relevant epochs potentially containing seizures or other paroxysmal episodes obtaining promising results.

For other video-based approaches in this field the reader is referred to [28, 30]. Unfortunately, these methods did not capture the complexity and the subtle differences of different types of SHE.

In this paper, we propose an easy-to-use and effective end-to-end pipeline based on computer vision and deep learning methods as a support for physicians in the differential diagnosis of semeiologic interpretation of SHE episodes and DOA. In particular, among the different deep learning architectures, we focused on 3D Convolutional Neural Networks (3D CNN) [37] for their ability to highlight both spatial and temporal information. Moreover, we addressed the multi-class classification by adopting a two-stage approach based on one-vs-one class classification. This allowed our architecture to focus on different features while comparing different pairs of classes.

2 | MATERIALS AND METHODS

2.1 | Dataset

The video-Polysomnographic (vPSG) recordings conducted for sleep-related paroxysmal events were part of a dataset acquired by the Sleep Medicine Center, "C. Munari" Center of Epilepsy Surgery, Niguarda Hospital, Milan, the Sleep Medicine Unit, Neurocenter of Southern Switzerland, Regional Hospital of Lugano,

Lugano, Switzerland and the Child Neuropsychiatry Unit, IRCCS G. Gaslini, Genoa, Italy. The current retrospective study received the approval of the Niguarda Hospital ethics committee (ID 939–12.12.2013). All patients gave written consent for the use of video recordings for research and publications purposes.

We considered only the video satisfying the following criteria: one major habitual episode recorded and people with definite diagnosis of either SHE or DOA [4, 8]. SHE episodes were divided into four clusters of SP:

- SP1: elementary motor signs;
- SP2: unnatural hypermotor movements;
- SP3: integrated hypermotor movements;
- SP4: gestural behaviors with high emotional content.

This categorization was performed by expert physicians and extensively described in [4] and it was adopted in this paper as ground truth for our deep learning pipeline. Every SP designation was reviewed among five epileptologists (L.N., L.T., P.P., S.A.G, G.N.) in order to reach an agreement in the event of discordant SP categorization.

To reduce imbalance, in this study, we considered the same number of videos for each class. The class with the lowest number of acquisitions was composed of 9 different people. Since the dataset had 5 classes (*i.e.,* the 4 different semiology patterns of SHE mentioned above and DOA), this resulted in a total of 45 videos, one for each participant (see Supplemental Table 1 for age and sex information about the participants).

INSERT Supplemental TABLE 1 HERE

The videos were analyzed retrospectively and were originally recorded to allow subsequent visual inspections by medical experts. For this reason, the videos have been acquired using different devices and without following a specific acquisition protocol and setup. As a consequence, they have a high variability in terms of spatial and temporal resolution, duration, different lighting conditions and various background settings. In particular, they had temporal resolutions varying between 25 and 30 frames per seconds (fps), average duration $\pm$ standard deviation of 70 $\pm$ 33 seconds and [minimum, maximum] duration of [13, 149] seconds. Figure 1 shows the setup necessary for the analysis and some examples of video-acquisition.

INSERT FIGURE 1 HERE

2.2 | Implemented deep learning pipeline

In this study, we designed and validated an automatic video analysis pipeline leveraging deep learning algorithms. Our approach was based on a set of binary classifiers able to discriminate between pairs of classes (*i.e*., four semiology patterns of SHE and DOA). Then, we merged the results of the binary classification with a custom multi-class architecture taking video inputs and automatically detecting the presence of different types of SHE or DOA episodes. In the remainder of the section, we will describe in detail the binary classification models, their training, validation (see Figure 2) and the subsequent merging procedure (see Figure 3).

2.2.1 | Binary classifiers training and validation

We designed one classifier for each possible pair of classes in the dataset, resulting in a total of 10 binary classifiers, each one learning to discriminate between two classes, specifically SP1 vs SP2, SP1 vs SP3, SP1 vs SP4, SP1 vs DOA, SP2 vs SP3, SP2 vs SP4, SP2 vs DOA, SP3 vs SP4, SP3 vs DOA, SP4 vs DOA.

As for the binary classification algorithm we considered a 3D Convolutional Neural Network (3D CNN): we implemented a custom version of the 3D deep learning architecture presented in [37] (*i.e.*, I3D model). The latter had been proven to be effective in several application domains related with video analysis [38, 39, 40]. We now report more details on the binary classification method, specifically on how it processed the input video. We split the video in temporal windows *Tw* composed by *n* frames each. Based on empirical considerations, we set n=25 frames. The obtained time windows were used as inputs for our classifiers. Thus, the designed model provided one label per each time window. The ground truth labels used for training and validation were determined by expert clinicians, following the study presented in [4] and they were associated with each time window. The predictions for each time window were then grouped to create a single predicted label for the entire video, referred to as the "video prediction" in Figure 2. To compute the final outcome of each video, following the standard procedure in machine learning studies, we considered the class predicted more frequently (more than 50% of the total amount of time windows extracted). Figure 2 shows a schematic description of this first part of our pipeline.

INSERT FIGURE 2 HERE

As for the training and validation of each binary classifier, we prepared the dataset according to the standard procedure usually adopted in supervised deep learning approaches: we divided the dataset in two groups, the *training set* and the *test set*. In supervised learning, training a deep learning architecture involves presenting the model with examples and adjusting the model's internal parameters to minimize the error between the model's predictions and the true labels of the examples (assigned by expert physicians during the dataset preparation) [41]. After training, the test set was used only to evaluate the model's performance on completely new data that the algorithm did not use during training. With this procedure, we assessed the generalization performance of the algorithm on previously unseen videos.

To this purpose, we built the training set by randomly selecting 7 out of 9 videos for each class (*i.e.*, 80% of their total amount), keeping the other 2 videos per class as the test set.

The training phase involved the validation of a given model, by means of a standard machine learning procedure called leave-one-out cross-validation. Leave-one-out cross-validation is usually adopted when the dataset is relatively small [42] to train classifiers robustly. In our case, it was carried out as follows: among the 7 training videos of each class, we selected 6 videos for the actual training and 1 video for the validation, and we repeated this selection for each possible left-out video resulting in 7 different classifiers. Therefore, in each training step, we considered 12 videos for training and 2 for validation.

2.2.2 | Multi-class merging procedure

Given a video, we set up a merging procedure that combined the results of each binary classifier and returned a single prediction for the input video. Specifically, for a given input video, each of the 10 binary classifiers provided a label, with a total of 10 labels ("binary classifiers video predictions" in Figure 3). The final outcome of the overall multi-class procedure was computed as the label recurring most frequently in the 10 output labels. Figure 3 provides a visual representation of this phase. If two or more classes had the same number of occurrences, we set up a procedure that considered the higher accuracy in the time windows classification to decide the final outcome. However, this scenario never happened in our analysis.

INSERT FIGURE 3 HERE

Here again, to test our procedure, we used 2 videos for each class, those that we excluded from the training, and validation procedure and we performed each step of the pipeline. In this way, we adopted 10 test videos previously unseen by the network.

3 | RESULTS

The results are presented as follows. First, we presented the results for the overall procedure (merging the predictions of the binary classifiers) in two videos per class previously unseen by the models (test set). Then, we reported the mean validation performance of the different binary classifiers in the classification of each time window, highlighting the potential of the proposed methods in the classification of each binary problem; in particular we focused on:

- accuracy: the total number of time windows correctly predicted for both classes;
- sensitivity: the total number of time windows correctly predicted for the first class;

Finally, we reported the performance of each binary classifier in the classification of each video, merging the predictions of the time windows.

We tested our overall procedure on the test set composed of 10 videos, considering the 2 videos per class randomly excluded from the training and validation analysis. Our algorithm was able to correctly classify 8 test videos out of 10 reaching a **test accuracy of 80%**. In particular, one of the two misclassified videos belonged to DOA (ground truth label) and it was classified as SP4. The other misclassified video belonged to SP3 and it was classified as SP2. For this reason we can also affirm that we reached 90% accuracy in the classification between epilepsy and DOA.

Table 1 reports the results of the 10 binary classifiers. Each row refers to the binary classifier applied on the pair of classes indicated in the first column. Column A reports the mean, the standard deviation of the validation (leave-one-out) accuracy in the classification of each time window *Tw* and the 95% Confidence Interval (C.I.). Column B reports the sensitivity values (mean and standard deviation). Column C reports the overall video accuracy reached by the binary classifiers obtained by combining the information of each time window *Tw* (*i.e.*, we considered as final prediction the class that occurred more than 50% of the total amount of time windows for each video). Results showed that both accuracy (Table 1, Column A) and sensitivity (Table 1, Column B) values were above 90% in all tests apart from the comparison between SP2 vs SP3 and SP3 vs SP4, where we highlighted both lower mean and higher standard deviation values. In these cases the % was computed considering the number of time windows correctly predicted with respect to the total number of time windows extracted for each video. In particular, considering the duration of the videos reported in Section 2, we had a minimum and maximum number of time windows respectively equal to 390 and 3725.

INSERT TABLE 1 HERE

As for the overall video accuracy (Table 1, Column C), 100 percent of accuracy was achieved for all but two tests (SP2 vs SP3, 86%; SP3 vs SP4, 86%). Here the % is referred to the number of videos correctly predicted over their total number considered for the training and the validation procedures (*i.e.*, 14 for each binary classifier).

4 | DISCUSSION

The results obtained in this paper showed the potential of the presented pipeline in the discrimination of different SP of SHE and DOA, highlighting the possibility to introduce and use automatic algorithms based on 3D CNN as a tool to support the differential diagnosis performed by expert physicians. In fact, we reached promising results (**80% of test accuracy**) in the discrimination of the four different SHE semiology patterns and DOA. This result is obtained considering 10 video examples previously unseen by the deep learning algorithm. This sample size is small and we plan to increase it in the future, but it is enough to highlight the potential of our procedure. The presented pipeline could also be adopted as a fast, non-invasive and easy to use way to highlight possible differences in motion patterns to be explored and deepened with more accurate eeg-based or marker-based techniques. In this case, we used the implemented pipeline to analyze Video-Polysomnography records acquired in the laboratory environment, but it can be adopted without any modification to home video-recordings.

At the beginning, we started by treating the problem as a multi-class classification problem: we trained a deep convolutional architecture (3D CNN) with all the 5 classes at the same time. In this way, given an input video, with only one deep learning model we should be able to classify among all the different classes in our dataset. Due to the complexity of the problem (*i.e.*, similar motion characteristics among different classes) and to the relatively small amount of data for this type of algorithms, we reached an overall mean accuracy during the test of 60%. For this reason, we implemented the procedure based on the combination of binary classifiers described and presented in this paper.

In terms of overall classification accuracy, the procedure presented here is in line with the results obtained by other studies. For example, Ahmedt-Aristizabal et al. [29] obtained an accuracy of 68.1% to support the identification of semiology patterns adopting the landmark-based approach enriched with optical flow information and an accuracy of 79.6% adopting the region-based approach (*i.e.*, the end-to-end pipeline that leveraged 2D CNN and LSTM).

However, with respect to other works, our method presents some advantages. In fact, since it started from the classification of different time windows, we could extract a measure related to the uncertainty of the prediction. For instance, if the number of time windows correctly detected was a high percentage of their total number we could be more confident about the final prediction with respect to a lower percentage. Also, starting from the time windows classification, it was possible to highlight different semiology patterns for the same video, increasing the potential of the implemented tool as it may highlight common patterns in different SP. In addition, it would be possible to consider and reason on different thresholds to assign the final video prediction (and not only the 50% threshold that we used in this work) to improve the overall accuracy.

Furthermore, starting from binary classifiers that compared pairs of classes, our pipeline allowed us to focus on the complexity of each binary problem. From Table 1, it is possible to appreciate that there are certain binary problems that were more difficult to perform and that were less stable (bigger standard deviation) than others. Interestingly, the two binary tasks that presented more uncertainty are: (a) the discrimination between SP2 vs SP3 and (b) between SP3 and SP4, the same that are difficult to discriminate also by expert physicians [4]. Indeed, there is a continuum between SP1 and SP4 [4], and, while the distinction between motor patterns belonging to SP1 (tonic/dystonic asymmetrical posture) and SP4 (hyperkinetic behavior with affective symptoms) can be easily made, a distinction between SP2 and SP3 and between SP3 and SP4 may not be so obvious or feasible.

Similarly, in the merging procedure after the training and validation analysis, the two videos that were misclassified belonged to an SP3 episode that was classified as SP2 and a DOA episode that was classified as SP4. Indeed, both DOA and SP4 may be characterized by hypermotor manifestations of various complexities, with bizarre and violent behaviors and their distinction may be challenging [10]. These findings suggest that the 3D CNNs may rely on considerations similar to those evaluated by physicians.

On the other hand, our pipeline is not interpretable, making it difficult to understand what are the specific characteristics that guided the classification (*i.e.*, black box algorithm). This can be further explored by adding interpretable tools [43] that will give an idea on the steps and on the portions of the video that guided the 3D CNN outcome.

5 | CONCLUSIONS

The preliminary results obtained in this study showed the potential of our procedure based on deep learning in the classification of SHE seizures. Moreover, this pipeline may represent a supporting tool for expert physicians in the final differential diagnosis among different SHE semiology patterns and DOA. Of course the work carried on in this paper needs to be further explored. In particular, future directions that should be considered to improve the overall procedure to make it more stable and reliable are: (i) to enlarge the population involved in the analysis (internal validity); this is particularly important for generalization purposes and to increase the number of test videos; (ii) to test our procedure in an independent external population (external validity); (iii) to deepen the meaning behind the time windows classification and explore the role of the threshold that is used to assign a final prediction to each video; (iv) to design a more accurate merging procedure; (v) to increase the interpretability (*i.e.*, try to understand the motivation behind the choice of the deep learning modules in the classification task) using techniques to understand the portions of the videos that influences the decision; (vi) to explore and deepen an approach that, starting from the estimation of the pose and the detection of interesting keypoints on the image plane, extract quantitative parameters that are representative of the motion patterns; (vii) to study and deepen the potential of optical flow algorithms.

Acknowledgement

This work was carried out within the framework of the project "RAISE - Robotics and AI for Socio-economic Empowerment" and has been supported by European Union - NextGenerationEU.

Work supported by #NEXTGENERATIONEU (NGEU) and funded by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), project MNESYS (PE0000006) – A Multiscale integrated approach to the study of the nervous system in health and disease (DN. 1553 11.10.2022).

Credit authorship contribution statement

Matteo Moro: Data analysis, Methodology, Software, Writing - Original draft preparation. **Vito Paolo Pastore**: Data analysis, Methodology, Software, Writing - Original draft preparation. **Giorgia Marchesi**: Data analysis, Writing - Review and editing. **Paola Proserpio**: Data curation, Writing - Review and editing. **Laura Tassi**: Data curation, Writing - Review and editing. **Anna Castelnovo**: Data curation, Writing - Review and editing. **Mauro Manconi**: Data curation, Writing - Review and editing. **Giulia Nobile**: Data curation, Writing - Review and editing. **Ramona Cordani**: Data curation, Writing - Review and editing. **Steve A. Gibbs:** Data curation, Writing - Review and editing. **Francesca Odone**: Conceptualization of the study, Methodology, Writing - Original draft preparation and Review and editing. **Maura Casadio**: Conceptualization of the study, Methodology, Writing - Original draft preparation and Review and editing. **Lino Nobili**: Conceptualization of the study, Data curation, Methodology, Writing - Original draft preparation and Review and editing.

References

[1] Tinuper P, Bisulli F, Cross J, Hesdorffer D, Kahane P, Nobili L, et al. Definition and diagnostic criteria of sleep-related hypermotor epilepsy. Neurology 2016;86(19):1834–1842.

[2] Mai R, Sartori I, Francione S, Tassi L, Castana L, Cardinale F, et al. Sleep-related hyperkinetic seizures: always a frontal onset? Neurological Sciences 2005;26(3):s220–s224.

[3] Proserpio P, Cossu M, Francione S, Tassi L, Mai R, Didato G, et al. Insular-opercular seizures manifesting with sleep-related paroxysmal motor behaviors: a stereo-EEG study. Epilepsia 2011;52(10):1781–1791.

[4] Gibbs SA, Proserpio P, Francione S, Mai R, Cardinale F, Sartori I, et al. Clinical features of sleep-related hypermotor epilepsy in relation to the seizure-onset zone: a review of 135 surgically treated cases. Epilepsia 2019;60(4):707–717.

[5] Castelnovo A, Lopez R, Proserpio P, Nobili L, Dauvilliers Y. NREM sleep parasomnias as disorders of sleep-state dissociation. Nature Reviews Neurology 2018;14(8):470–481.

[6] Derry CP, Harvey AS, Walker MC, Duncan JS, Berkovic SF. NREM arousal parasomnias and their distinction from nocturnal frontal lobe epilepsy: a video EEG analysis. Sleep 2009;32(12):1637–1644.

[7] Derry CP, Davey M, Johns M, Kron K, Glencross D, Marini C, et al. Distinguishing sleep disorders from seizures: diagnosing bumps in the night. Archives of neurology 2006;63(5):705–709.

[8] Proserpio P, Loddo G, Zubler F, Ferini-Strambi L, Licchetta L, Bisulli F, et al. Polysomnographic features differentiating disorder of arousals from sleep-related hypermotor epilepsy. Sleep 2019;42(12):zsz166.

[9] Loddo G, Baldassarri L, Zenesini C, Licchetta L, Bisulli F, Cirignotta F, et al. Seizures with paroxysmal arousals in sleep-related hypermotor epilepsy (SHE): Dissecting epilepsy from NREM parasomnias. Epilepsia 2020;61(10):2194–2202.

[10] Montini A, Loddo G, Baldelli L, Cilea R, Provini F. Sleep-related hypermotor epilepsy vs disorders of arousal in adults: a step-wise approach to diagnosis. Chest 2021;160(1):319–329.

[11] Nijsen TM, Arends JB, Griep PA, Cluitmans PJ. The potential value of three-dimensional accelerometry for detection of motor seizures in severe epilepsy. Epilepsy & Behavior 2005;7(1):74–84.

[12] Gubbi J, Kusmakar S, Rao AS, Yan B, O'Brien T, Palaniswami M. Automatic detection and classification of convulsive psychogenic nonepileptic seizures using a wearable device. IEEE journal of biomedical and health informatics 2015;20(4):1061–1072.

[13] Becq G, Kahane P, Minotti L, Bonnet S, Guillemaud R. Classification of epileptic motor manifestations and detection of tonic–clonic seizures with acceleration norm entropy. IEEE Transactions on Biomedical Engineering 2013;60(8):2080–2088.

[14] Cuppens K, Chen CW, Wong KBY, Van de Vel A, Lagae L, Ceulemans B, et al. Integrating video and accelerometer signals for nocturnal epileptic seizure detection. In: Proceedings of the 14th ACM international conference on Multimodal interaction; 2012. p. 161–164.

[15] Patel S, Lorincz K, Hughes R, Huggins N, Growdon J, Standaert D, et al. Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors. IEEE transactions on information technology in biomedicine 2009;13(6):864–873.

[16] Lopez-Nava IH, Munoz-Melendez A. Wearable inertial sensors for human motion analysis: A review. IEEE Sensors Journal 2016;16(22):7821–7834.

[17] Chen L, Yang X, Liu Y, Zeng D, Tang Y, Yan B, et al. Quantitative and trajectory analysis of movement trajectories in supplementary motor area seizures of frontal lobe epilepsy. Epilepsy & Behavior 2009;14(2):344–353.

[18] Carse B, Meadows B, Bowers R, Rowe P. Affordable clinical gait analysis: An assessment of the marker tracking accuracy of a new low-cost optical 3D motion analysis system. Physiotherapy 2013;99(4):347–351.

[19] Cunha JPS, Choupina HMP, Rocha AP, Fernandes JM, Achilles F, Loesch AM, et al. NeuroKinect: a novel low-cost 3Dvideo-EEG system for epileptic seizure motion quantification. PloS one 2016;11(1):e0145669.

[20] Colyer SL, Evans M, Cosker DP, Salo AI. A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system. Sports medicine-open 2018;4(1):1–15.

[21] Needham L, Evans M, Cosker DP, Wade L, McGuigan PM, Bilzon JL, et al. The accuracy of several pose estimation methods for 3D joint centre localisation. Scientific reports 2021;11(1):1–11.

[22] Needham L, Evans M, Cosker DP, Colyer SL. Can markerless pose estimation algorithms estimate 3D mass centre positions and velocities during linear sprinting activities? Sensors 2021;21(8):2889.

[23] Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer vision: A brief review. Computational intelligence and neuroscience 2018;2018.

[24] Pastore VP, Moro M, Odone F. A semi-automatic toolbox for markerless effective semantic feature extraction. Scientific Reports 2022;12(1):1–12.

[25] Cao Z, Simon T, Wei SE, Sheikh Y. Realtime multi-person 2d pose estimation using part affinity fields. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 7291–7299.

[26] Moro M, Marchesi G, Odone F, Casadio M. Markerless gait analysis in stroke survivors based on computer vision and deep learning: A pilot study. In: Proceedings of the 35th annual ACM symposium on applied computing; 2020. p. 2097–2104.

[27] Moro M, Pastore VP, Tacchino C, Durand P, Blanchi I, Moretti P, et al. A markerless pipeline to analyze spontaneous movements of preterm infants. Computer Methods and Programs in Biomedicine 2022;226:107119.

[28] 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):1817–1831.

[29] Ahmedt-Aristizabal D, Denman S, Nguyen K, Sridharan S, Dionisio S, Fookes C. Understanding patients' behavior: Vision based analysis of seizure disorders. IEEE journal of biomedical and health informatics 2019;23(6):2583–2591.

[30] Abbasi B, Goldenholz DM. Machine learning applications in epilepsy. Epilepsia 2019;60(10):2037–2047.

[31] Karácsony T, Loesch-Biffar AM, Vollmar C, Rémi J, Noachtar S, Cunha JPS. Novel 3D video action recognition deep learning approach for near real time epileptic seizure classification. Scientific Reports 2022;12(1):1–13.

[32] Brox T, Bruhn A, Papenberg N, Weickert J. High accuracy optical flow estimation based on a theory for warping. In: European conference on computer vision Springer; 2004. p. 25–36.

[33] Pathak D, Girshick R, Dollár P, Darrell T, Hariharan B. Learning features by watching objects move. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 2701–2710.

[34] Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation 1997;9(8):1735–1780.

[35] Peltola J, Basnyat P, Armand Larsen S, Østerkjærhuus T, Vinding Merinder T, Terney D, Beniczky S. Semiautomated classification of nocturnal seizures using video recordings. Epilepsia 2022;00:1–7.

[36] Armand Larsen S, Terney D, Østerkjerhuus T, Vinding Merinder T, Annala K, Knight A, Beniczky S. Automated detection of nocturnal motor seizures using an audio-video system. Brain and Behavior 2022;12(9), e2737.

[37] Carreira J, Zisserman A. Quo vadis, action recognition? a new model and the kinetics dataset. In: proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017. p. 6299–6308.

[38] Chen CFR, Panda R, Ramakrishnan K, Feris R, Cohn J, Oliva A, et al. Deep analysis of cnn-based spatio-temporal representations for action recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2021. p. 6165–6175.

[39] Kocabas M, Athanasiou N, Black MJ. Vibe: Video inference for human body pose and shape estimation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2020. p. 5253–5263.

[40] Zhang HB, Zhang YX, Zhong B, Lei Q, Yang L, Du JX, et al. A comprehensive survey of vision-based human action recognition methods. Sensors 2019;19(5):1005.

[41] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436–444.

[42] Hastie T, Tibshirani R, Friedman JH, Friedman JH. The elements of statistical learning: data mining, inference, and prediction, vol. 2. Springer; 2009.

[43] Augasta MG, Kathirvalavakumar T. Reverse engineering the neural networks for rule extraction in classification problems. Neural processing letters 2012;35(2):131–150.

FIGURE AND TABLES LEGENDS

Figure 1. Setup (left panel) and examples of frames of different videos (right panel).

Figure 2. Video classification with one binary classifier. The video is first divided in different time windows *Tw*. Each time window is then classified. The predictions for the *Tw* of the same video are grouped using the 50% threshold described above and a single prediction for the whole video is provided ("video prediction").

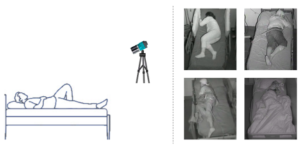
Figure 3. Testing procedure - combination of binary classifiers. Starting from a video, we reach a final outcome considering the outputs of each binary classifier. Each binary block is composed as described in Figure 2.

TABLE 1. Binary classifiers' performance. For each row, we report the mean, the standard deviation of the performance of each binary classifier (with leave-one-out cross-validation described in Section 2) and the 95% Confidence Interval (C.I.). SP: semiology pattern; DOA: disorders of Arousal.

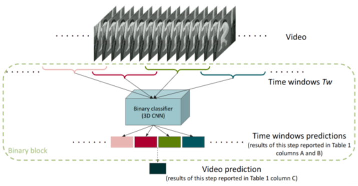
Supplemental TABLE 1. Participants information (gender and age). SP: semiology pattern; DOA: Disorder of Arousal.

TABLE 1. Binary classifiers' performance. For each row, we report the mean, the standard deviation of the performance of each binary classifier (with leave-one-out cross-validation described in Section 2) and the 95% Confidence Interval (C.I.). SP: semiology pattern; DOA: disorders of Arousal.

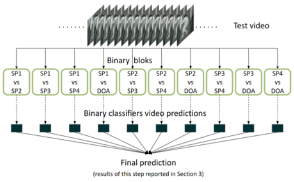
	A	B	C
Classes	**Mean accuracy (standard deviation) [C.I.] -% time windows**	**Mean sensitivity (standard deviation) -% time windows**	**Accuracy - whole videos %**
SP1 vs SP2	95% (3%) [93.9 - 96.1]	93% (3%)	100%
SP1 vs SP3	96% (3%) [94.9 - 97.1]	95% (4%)	100%
SP1 vs SP4	94% (5%) [92.2 - 95.8]	93% (3%)	100%
SP1 vs DOA	97% (2%) [96.3 - 97.7]	98% (2%)	100%
SP2 vs SP3	87% (15%) [81.6 - 92.4]	81% (17%)	86%
SP2 vs SP4	93% (3%) [91.9 - 94.1]	94% (2%)	100%
SP2 vs DOA	93% (4%) [91.6 - 94.4]	91% (3%)	100%
SP3 vs SP4	82% (29%) [71.6 - 92.4]	84% (20%)	86%
SP3 vs DOA	94% (3%) [92.9 - 95.1]	95% (4%)	100%
SP4 vs DOA	93% (7%) [90.5 - 95.5]	92% (3%)	100%



EPI_17605_Moro-fig1.tif



EPI_17605_Moro-fig2.tif



EPI_17605_Moro-fig3.tif