Privacy-Preserving Early Detection of Epileptic Seizures in Videos

Deval Mehta[1,2,3](✉), Shobi Sivathamboo[4,5,6], Hugh Simpson[4,5],
Patrick Kwan[4,5,6], Terence O'Brien[4,5], and Zongyuan Ge[1,2,3,7]

[1] AIM for Health Lab, Faculty of IT, Monash University, Melbourne, Australia
deval.mehta@monash.edu
[2] Monash Medical AI, Monash University, Melbourne, Australia
[3] Faculty of Engineering, Monash University, Melbourne, Australia
[4] Department of Neuroscience, Central Clinical School, Faculty of Medicine Nursing and Health Sciences, Monash University, Melbourne, Australia
[5] Department of Neurology, Alfred Health, Melbourne, Australia
[6] Departments of Medicine and Neurology, The University of Melbourne, Royal Melbourne Hospital, Parkville, VIC, Australia
[7] Airdoc-Monash Research Lab, Monash University, Melbourne, Australia
https://www.monash.edu/it/aimh-lab/home

Abstract. In this work, we contribute towards the development of video-based epileptic seizure classification by introducing a novel framework (SETR-PKD), which could achieve privacy-preserved early detection of seizures in videos. Specifically, our framework has two significant components - (1) It is built upon optical flow features extracted from the video of a seizure, which encodes the seizure motion semiotics while preserving the privacy of the patient; (2) It utilizes a transformer based progressive knowledge distillation, where the knowledge is gradually distilled from networks trained on a longer portion of video samples to the ones which will operate on shorter portions. Thus, our proposed framework addresses the limitations of the current approaches which compromise the privacy of the patients by directly operating on the RGB video of a seizure as well as impede real-time detection of a seizure by utilizing the full video sample to make a prediction. Our SETR-PKD framework could detect tonic-clonic seizures (TCSs) in a privacy-preserving manner with an accuracy of **83.9%** while they are only **half-way** into their progression. Our data and code is available at https://github.com/DevD1092/seizure-detection.

Keywords: epilepsy · early detection · knowledge distillation

1 Introduction

Epilepsy is a chronic neurological condition that affects more than 60 million people worldwide in which patients experience epileptic seizures due to abnormal

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43904-9_21.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14224, pp. 210–219, 2023.
https://doi.org/10.1007/978-3-031-43904-9_21

brain activity [17]. Different types of seizures are associated with the specific part of the brain involved in the abnormal activity [8]. Thus, accurate detection of the type of epileptic seizure is essential to epilepsy diagnosis, prognosis, drug selection and treatment. Concurrently, real-time seizure alerts are also essential for caregivers to prevent potential complications, such as related injuries and accidents, that may result from seizures. Particularly, patients suffering from tonic-clonic seizures (TCSs) are at a high risk of sudden unexpected death in epilepsy (SUDEP) [18]. Studies have shown that SUDEP is caused by severe alteration of cardiac activity actuated by TCS, leading to immediate death or cardiac arrest within minutes after the seizure [5]. Therefore, it is critical to accurately and promptly detect and classify epileptic seizures to provide better patient care and prevent any potentially catastrophic events.

The current gold standard practice for detection and classification of epileptic seizures is the hospital-based Video EEG Monitoring (VEM) units [23]. However, this approach is expensive and time consuming which is only available at specialized centers [3]. To address this issue, the research community has developed automated methods to detect and classify seizures based on several modalities - EEG [7,30], accelerometer [16], and even functional neuroimaging modalities such as fMRI [22] and electrocorticography (ECoG) [24]. Although, there have been developments of approaches for the above modalities, seizure detection using videos remains highly desirable as it involves no contact with the patient and is easier to setup and acquire data compared to other modalities. Thus, researchers have also developed automated approaches for the video modality.

Initial works primarily employed hand-crafted features based on patient motion trajectory by attaching infrared reflective markers to specific body key points [4,15]. However, these approaches were limited in performance due to their inability to generalize to changing luminance (night time seizures) or when the patient is occluded (covered by a bed sheet) [14]. Thus, very recently deep learning (DL) models have been explored for this task [1,2,12,21,29]. [29] demonstrated that DL models could detect generalized tonic-clonic seizures (GTCSs) from the RGB video of seizures. Authors in [21] radically used transfer learning (from action recognition task) to train DL networks for distinguishing focal onset seizures (FOSs) from bilateral TCSs using features extracted from the RGB video of seizures. Whereas, the authors in [12] developed a DL model to discriminate dystonia and emotion in videos of Hyperkinetic seizures. However, these developed approaches have two crucial limitations - (1) As these approaches directly operate on RGB videos, there is a possibility of privacy leakage of the sensitive patient data from videos. Moreover, obtaining consent from patients to share their raw RGB video data for building inter-cohort validation studies and generalizing these approaches on a large scale becomes challenging; (2) The current approaches consider the full video of a seizure to make predictions, which makes early detection of seizures impossible. The duration of a seizure varies significantly among patients, with some lasting as short as 30 s while others can take minutes to self-terminate. Thus, it is unrealistic to wait until the completion of a long seizure to make a prediction and alert caregivers.

In this work, we address the above two challenges by building an in-house dataset of privacy-preserved extracted features from a video and propose a framework for early detection of seizures. Specifically, we investigate two aspects - (1) The feasibility of detecting and classifying seizures based only on *optical flow*, a modality that captures temporal differences in a scene while being intrinsically privacy-preserving. (2) The potential of predicting the type of seizure during its progression by analyzing only a fraction of the video sample. Our early detection approach is inspired by recent developments in early action recognition in videos [9,10,19,21,28,31]. We develop a custom feature extractor-transformer framework, named **SE**izure **TR**ansformer (SETR) block for processing a single video sample. To achieve early detection from a fraction of the sample, we propose **P**rogressive **K**nowledge **D**istillation (PKD), where we gradually distill knowledge from SETR blocks trained on longer portions of a video sample to SETR blocks which will operate on shorter portions. We evaluate our proposed SETR-PKD framework on two datasets - an in-house dataset collected from a VEM unit in a hospital and a publicly available dataset of video-extracted features (GESTURES) [21]. Our experiments demonstrate that our proposed SETR-PKD framework can detect TCS seizures with an accuracy of **83.9%** in a privacy-preserving manner when they are only **half-way** into their progression. Furthermore, we comprehensively compare the performance of direct knowledge distillation with our PKD approach on both optical flow features (in-house dataset) and raw video features (public dataset). We firmly believe that our proposed method makes the first step towards developing a privacy-preserving real-time system for seizure detection in clinical practice.

2 Proposed Method

In this section, we first outline the process of extracting privacy-preserving information from RGB video samples to build our in-house dataset. Later, we explain our proposed approach for early detection of seizures in a sample.

2.1 Privacy Preserving Optical Flow Acquisition

Our in-house dataset of RGB videos of patients experiencing seizures resides on hospital premises and is not exportable due to the hospital's ethics agreement[1]. To work around this limitation, we develop a pipeline to extract optical flow information [11] from the videos. This pipeline runs locally within the hospital and preserves the privacy of the patients while providing us with motion semiotics of the seizures. An example of the extracted optical flow video sample can be seen in Fig. 1. We use the TV-L1 algorithm [20] to extract the optical flow features for each video, which we then export out of the hospital for building our proposed approach. We provide more information about our dataset, including the number of patients and seizures, annotation protocol, etc. in Sect. 3.

[1] We have a data ethics agreement approved for collection of data at hospital.

2.2 Early Detection of Seizures in a Sample

Consider an input optical flow video sample V_i as shown in Fig. 1(a) with a time period of T_i, consisting of N frames - $\{f_0, f_1, ...f_{N-1}\}$, and having a ground truth label of $y_i \in \{0, 1, ...C\}$ where is C the total number of categories. Then, the task of early detection is to build a framework that could classify the category of the sample correctly by analyzing the least possible partial segment of the sample. Thus, to define the problem of early detection, we split the sample V_i into k segments -$\{0, 1, ...k-1\}$ starting from the beginning to the end as shown in Fig. 1(b). Here V_i^{k-1} corresponds to the full video sample and the descending segments correspond to the reduced partial video samples. We build these partial segments by equally adding the temporal information throughout the sample i.e. the time period for a partial subset V_i^j of a sample V_i is computed as $(j+1) \times T_i/k$. Thus, the early detection task is to correctly predict the category y_i of the sample V_i from the lowest possible (j) partial segment V_i^j of V_i. In Fig. 1, we illustrate our proposed framework where - (a) First, we build a Seizure Transformer (SETR) block for processing a single optical flow video sample (b) Later, we employ SETR based Progressive Knowledge Distillation (SETR-PKD) to achieve early detection in a sample.

Processing a Single Sample. Since seizure patterns comprise of body movements, we implement transfer learning from a feature extractor pre-trained on action recognition task to extract the spatial features from the optical flow frames. Prior work [21] has shown that Temporal Segment Networks (TSNs) [27] pretrained on RGB videos of various actions are effective at extracting features from videos of seizures. We also utilize TSNs but pretrained on the optical flow modality, since we have privacy-preserved optical flow frames. The TSNs extract a 1D feature sequence for each frame f_j, referred as spatial features in Fig. 1(a). The spatial features are then processed by a linear transformation (1-layer MLP) that maps them into $motion_{tokens} \in \mathbb{R}^{N \times D}$, where each token has D-dimensions.

We leverage transformers to effectively learn temporal relations between the extracted spatial features of the seizure patterns. Following the strategy of ViT [6], after extracting the spatial features, we append a trainable class embedding $class_{embed} \in \mathbb{R}^D$ to the motion tokens. This class embedding serves to represent the temporal relationships between the motion tokens and is later used for classification ($class_{token}$ in Fig. 1(a)). As the order of the $motion_{tokens}$ is not known, we also add a learnable positional encoding $L_{POS} \in \mathbb{R}^{(N+1) \times D}$ to the combined $motion_{tokens}$ and $class_{embed}$. This is achieved using an element-wise addition and we term it as the input X_i for the input sample V_i.

To enable the interaction between tokens and learn temporal relationships for input sample classification, we employ the Vanilla Multi-Head Self Attention (MHSA) mechanism [26]. First, we normalize the input sequence $X_i \in \mathbb{R}^{(N+1) \times D}$ by passing it through a layer normalization, yielding $X_i^{'}$. We then use projection matrices $(Q_i, K_i, V_i) = (X_i^{'} W_i^Q, X_i^{'} W_i^K, X_i^{'} W_i^V)$ to project $X_i^{'}$ into queries (Q), keys (K), and values (V), where $W_i^{Q/K/V} \in \mathbb{R}^{D \times D}$ are the projection matrices

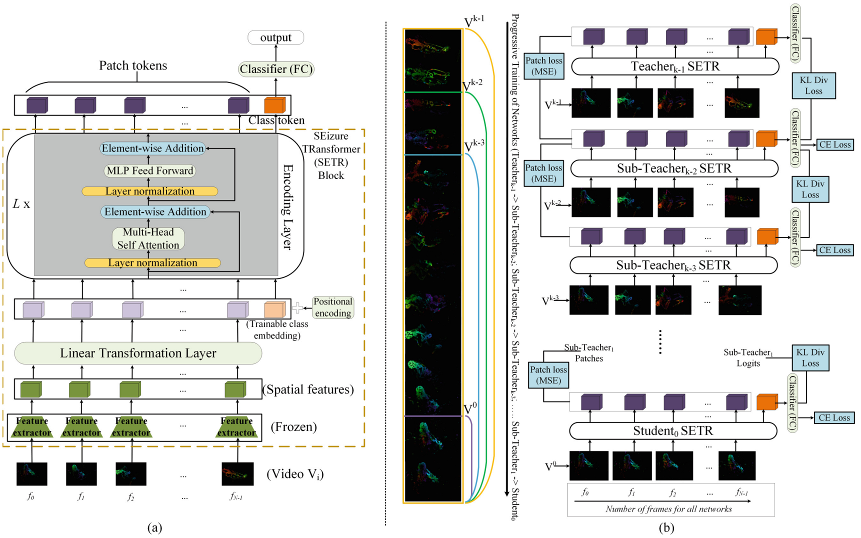

Fig. 1. Our proposed framework - (a) SEizure TRansformer (SETR) block for a single optical flow video sample (b) SETR based Progressive Knowledge Distillation (SETR-PKD) for early detection of seizures in a sample. (Best viewed in zoom and color).

for query, key, and value respectively. Next, we compute a dot product of Q with K and apply a softmax layer to obtain weights on the values. We repeat this self-attention computation N_h times, where N_h is the number of heads, and concatenate their outputs. Eq. 1, 2 depict the MHSA process in general.

$$A_i = Softmax(Q_i K_i) \tag{1}$$

$$MHSA(X_i^{'}) = A_i \times W_i^V, \qquad X_i^{'} = Norm(X_i) \tag{2}$$

Subsequently, the output of MHSA is passed to a two-layered MLP with GELU non-linearity while applying layer normalization and residual connections concurrently. Eq. 3, 4 represent this overall process.

$$m_l^{'} = MHSA(X_{l-1}^{'}) + X_{l-1}, \qquad l = 1...L \tag{3}$$

$$m_l = MLP(Norm(m_l^{'})) + m_l^{'}, \qquad l = 1...L \tag{4}$$

where $m_L \in \mathbb{R}^{(N+1)\times D}$ are the final output feature representations and L is the total number of encoding layers in the Transformer Encoder. Note that the first $\mathbb{R}^{N\times D}$ features correspond to the $patch_{tokens}$, while the final $\mathbb{R}^D$ correspond to the $class_{token}$ of the m_L as shown in Fig. 1(a). As mentioned earlier, we then use a one-layer MLP to predict the class label from the $class_{token}$. We refer to this whole process as a SEizure TRansformer (SETR) block shown in Fig. 1(a).

Progressive Knowledge Distillation. To achieve early detection, we use **K**nowledge **D**istillation in a **P**rogressive manner (PKD), starting from a SETR block trained on a full video sample and gradually moving to a SETR block trained on a partial video sample, as shown in Fig. 1(b). Directly distilling from a SETR block which has seen a significantly longer portion of the video (say V_i^{k-1}) to a SETR block which has only seen a smaller portion of the video sample (say V_i^0) will lead to considerable mismatches between the features extracted from the two SETRs as there is a large portion of the input sample that the $student_0$ SETR has not seen. In contrast, our proposed PKD operates in steps. First we pass the knowledge from teacher ($Teacher_{k-1}$ in Fig. 1(b)) SETR trained on V_i^{k-1} to a student ($Sub-teacher_{k-2}$) SETR that operates on V_i^{k-2}; Later, the $Sub-teacher_{k-2}$ SETR passes its distilled knowledge to its subsequent student ($Sub-teacher_{k-3}$) SETR, and this continues until the final $Sub-teacher_1$ SETR passes its knowledge to the bottom most $Student_0$ SETR. Since the consecutive segments of the videos do not differ significantly, PKD is more effective than direct distillation, which is proven by results in Sect. 3.4.

For distilling knowledge we consider both class token and patch tokens of the teacher and student networks. A standard Kullback-Leibler divergence ($\mathcal{L}_{KL}$) loss is applied between the probabilities generated from class token of the teacher and student SETR, whereas a mean squared error ($\mathcal{L}_{MSE}$) loss is computed between the patch tokens of teacher and student SETR. Overall, a student SETR is trained with three losses - $\mathcal{L}_{KL}$ and $\mathcal{L}_{MSE}$ loss for knowledge distillation, and a cross-entropy ($\mathcal{L}_{CE}$) loss for classification, given by the equations below.

$$\mathcal{L}_{KL} = \tau^2 \sum_j q_j^T (log(q_j^T / q_j^S)) \tag{5}$$

where q_j^S and q_j^T are the soft probabilities (moderated by temperature τ) of the student and teacher SETRs for the j^{th} class, respectively.

$$\mathcal{L}_{mse} = (\sum_{i=0}^{N} |p_i^T - p_i^S\|^2)/N \tag{6}$$

where N is the number of patches and p_i^T and p_i^S are the patches of teacher and student SETRs respectively.

$$\mathcal{L}_{total} = \mathcal{L}_{CE} + \alpha\mathcal{L}_{KL} + \beta\mathcal{L}_{mse} \tag{7}$$

where α and β are the weights for $\mathcal{L}_{KL}$ and $\mathcal{L}_{MSE}$ loss respectively.

3 Datasets and Experimental Results

3.1 In-House and Public Dataset

Our in-house dataset[2] contains optical flow information extracted from high-definition (1920 × 1080 pixels at 30 frames per second) video recordings of TCS

[2] We plan to release the in-house optical flow dataset and corresponding code.

seizures (infrared cameras are used for nighttime seizures) in a VEM unit in hospital. To annotate the dataset, two neurologists examined both the video and corresponding EEG to identify the clinical seizure onset (t_{ON}) and clinical seizure offset (t_{OFF}) times for each seizure sample. We curated a dataset comprising of 40 TCSs from 40 epileptic patients, with one sample per patient. The duration (in seconds) of the 40 TCSs in our dataset ranges from 52 to 367 s, with a median duration of 114 s. We also prepared normal samples (no seizure) for each patient by considering the pre-ictal duration from (t_{ON} - 300) to (t_{ON} - 60) seconds, resulting in dataset of 80 samples (40 normal and 40 TCSs). We refrain from using the 60 s prior to clinical onset as it corresponds to the transition period to the seizure containing preictal activity [13,25]. We use a 5-fold cross validation (split based on patients) for training and testing on our dataset.

We also evaluate the effectiveness of our early detection approach on the GESTURES dataset [21], which contains features extracted from RGB video samples of seizures. The dataset includes two seizure types - 106 focal onset seizures (FOS) and 77 Tonic-Clonic Seizures (TCS). In contrast to our in-house dataset, the features are provided by the authors, and we directly input them into our SETR block without using a feature extractor. To evaluate our method, we adopt the stratified 10-fold cross-validation protocol as used in GESTURES.

3.2 Training Implementation and Evaluation Metrics

We implement all experiments in PyTorch 1.8.1 on a single A100 GPU. The SETR block takes in a total of 64 frames (N) with 512 1-D spatial feature per frame, has 8 MHSA heads (N_h) with a dropout rate of 0.1, 3 encoder layers (L), and 256 hidden dimensions (D). For early detection, we experiment by progressively segmenting a sample into -{4,8,16} parts (k). We employ a grid search to select the weight of 0.2 and 0.5 for KL divergence (τ = 10) and MSE loss respectively. We train all methods with a batch size of 16, a learning rate of 1e-3 and use the AdamW optimizer with a weight decay of 1e-4 for a total 50 epochs. For GESTURES dataset, we implement a weighted BCE loss to deal with the dataset imbalance, whereas for our in-house dataset we implement the standard BCE loss. We use precision, recall and f1-score for benchmarking.

3.3 Performance for Early Detection

Table 1 shows the benchmarking performance of all techniques with varying fractions of input video samples on both datasets. We observed three key findings from the results in Table 1. First, transformer-based methods such as our proposed **SETR-PKD** and OaDTR exhibit better performance retention compared to LSTM-based techniques (RULSTM, Slowfast RULSTM, EgoAKD, GESTURES) with a reduction in the fraction of input sample. Second, **SETR-PKD** performance increases with k=8 from k=4, but saturates at k=16 for in-house dataset, whereas it achieves the best performance for k=4 for GESTURES dataset. The median seizure length for the in-house dataset and GESTURES dataset is 114 s and 71 s, respectively. As a result, PKD using relatively longer partial segments (k=4) is sufficient for GESTURES, while shorter partial segments (k=8) are required for our dataset. Thus, the optimal value of k for PKD may vary depending on a dataset. Finally, we observed better performance on the GESTURES dataset, which is expected given the more detailed and refined features extracted from RGB video compared to optical flow information.

3.4 Progressive V/s Direct Knowledge Distillation

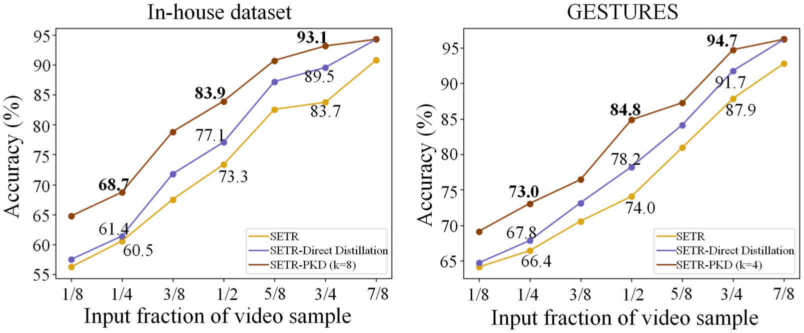

Fig. 2. Performance comparison of direct knowledge distillation and progressive knowledge distillation between SETR blocks for different fractions of input video sample.

To validate our approach of progressive knowledge distillation in a fair manner, we conducted an ablation study to compare it with direct knowledge distillation. Figure 2 shows the comparison of the accuracy of the two approaches for different fractions of the input video sample on both datasets. The results indicate that although direct knowledge distillation can increase performance, it is less effective when the knowledge gap is wide, i.e., from a SETR block trained on a full

Table 1. Benchmarking of different techniques for different fraction **{1/4, 1/2, 3/4, Full}** of input video sample. The performance is presented as mean of - **{Precision/Recall/F1-score}** across the 5-folds & 10-folds for in-house and GESTURES dataset respectively. (Best viewed in zoom).

Method/Dataset	In-house dataset				GESTURES			
	1/4	1/2	3/4	Full	1/4	1/2	3/4	Full
RULSTM [9]	0.57/0.56/0.56	0.72/0.71/0.71	0.79/0.79/0.79	0.95/0.93/0.94	0.65/0.64/0.64	0.71/0.73/0.72	0.84/0.85/0.84	0.93/0.94/0.93
Slowfast RULSTM [19]	0.57/0.56/0.56	0.73/0.72/0.72	0.81/0.80/0.80	0.94/0.94/0.94	0.67/0.65/0.66	0.73/0.72/0.72	0.86/0.84/0.85	0.97/0.95/0.96
EgoAKD [31]	0.64/0.65/0.64	0.79/0.80/0.79	0.89/0.90/0.89	0.95/0.94/0.94	0.70/0.69/0.69	0.80/0.79/0.79	0.93/0.90/91	0.97/0.94/0.95
OaDTR [28]	0.66/0.65/0.65	0.82/0.83/0.82	0.90/0.90/0.90	0.95/0.95/0.95	0.72/0.69/0.70	0.82/0.83/0.82	0.91/0.92/0.91	0.99/0.99/0.99
GESTURES [21]	0.59/0.60/0.59	0.74/0.73/0.73	0.82/0.83/0.82	0.94/0.94/0.94	0.68/0.66/0.66	0.74/0.72/0.73	0.86/0.85/0.85	0.97/0.99/0.98
SETR	0.61/0.60/0.60	0.75/0.73/0.74	0.84/0.83/0.83	0.96/0.95/0.95	0.67/0.66/0.66	0.73/0.74/0.73	0.88/0.88/0.88	0.98/0.99/0.98
SETR-PKD (k=4)	0.63/0.62/0.62	0.78/0.79/0.78	0.89/0.90/0.89	0.96/0.95/0.95	0.74/0.73/0.73	0.86/0.85/0.85	0.96/0.95/0.95	0.98/0.99/0.98
SETR-PKD (k=8)	0.70/0.69/0.69	0.86/0.84/0.85	0.92/0.93/0.92	0.96/0.95/0.95	0.73/0.74/0.73	0.85/0.85/0.85	0.95/0.96/0.95	0.98/0.99/0.98
SETR-PKD (k=16)	0.69/0.69/0.69	0.85/0.84/0.84	0.92/0.92/0.92	0.96/0.95/0.95	0.72/0.73/0.72	0.85/0.84/0.84	0.96/0.95/0.95	0.98/0.99/0.98

input sample to a SETR block trained on a minimal fraction of the input sample (1/8, 1/4, .. 1/2) compared to when the knowledge gap is small (5/8, .. 7/8). On the other hand, our SETR-PKD approach significantly improves performance for minimal fractions of input samples on both datasets.

4 Conclusion

In this work, we show that it is possible to detect epileptic seizures from optical flow modality in a privacy-preserving manner. Moreover, to achieve real-time seizure detection, we specifically develop a novel approach using progressive knowledge distillation which proves to detect seizures more accurately during their progression itself. We believe that our proposed privacy-preserving early detection of seizures will inspire the research community to pursue real-time seizure detection in videos as well as facilitate inter-cohort studies.

References

1. Ahmedt-Aristizabal, D., et al.: A hierarchical multimodal system for motion analysis in patients with epilepsy. Epilepsy Behav. **87**, 46–58 (2018)
2. Ahmedt-Aristizabal, D., Nguyen, K., Denman, S., Sridharan, S., Dionisio, S., Fookes, C.: Deep motion analysis for epileptic seizure classification. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 3578–3581. IEEE (2018)
3. Cascino, G.D.: Video-EEG monitoring in adults. Epilepsia **43**, 80–93 (2002)
4. Cunha, J.P.S., et al.: NeuroKinect: a novel low-cost 3Dvideo-EEG system for epileptic seizure motion quantification. PloS one **11**(1), e0145669 (2016)
5. Devinsky, O., Hesdorffer, D.C., Thurman, D.J., Lhatoo, S., Richerson, G.: Sudden unexpected death in epilepsy: epidemiology, mechanisms, and prevention. Lancet Neurol. **15**(10), 1075–1088 (2016)
6. Dosovitskiy, A., et al.: An image is worth 16×16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
7. Fan, M., Chou, C.A.: Detecting abnormal pattern of epileptic seizures via temporal synchronization of EEG signals. IEEE Trans. Biomed. Eng. **66**(3), 601–608 (2018)
8. Fisher, R.S., et al.: Operational classification of seizure types by the international league against epilepsy: position paper of the ILAE commission for classification and terminology. Epilepsia **58**(4), 522–530 (2017)
9. Furnari, A., Farinella, G.M.: Rolling-unrolling LSTMs for action anticipation from first-person video. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) **43**, 4021–4036 (2020)
10. Guan, W., et al.: Egocentric early action prediction via multimodal transformer-based dual action prediction. IEEE Trans. Circ. Syst. Video Technol. **33**(9), 4472–4483 (2023)
11. Horn, B.K., Schunck, B.G.: Determining optical flow. Artif. Intell. **17**(1–3), 185–203 (1981)
12. Hou, J.C., Thonnat, M., Bartolomei, F., McGonigal, A.: Automated video analysis of emotion and dystonia in epileptic seizures. Epilepsy Res. **184**, 106953 (2022)
13. Huberfeld, G., et al.: Glutamatergic pre-ictal discharges emerge at the transition to seizure in human epilepsy. Nat. Neurosci. **14**(5), 627–634 (2011)

14. Kalitzin, S., Petkov, G., Velis, D., Vledder, B., da Silva, F.L.: Automatic segmentation of episodes containing epileptic clonic seizures in video sequences. IEEE Trans. Biomed. Eng. **59**(12), 3379–3385 (2012)
15. Karayiannis, N.B., Tao, G., Frost, J.D., Jr., Wise, M.S., Hrachovy, R.A., Mizrahi, E.M.: Automated detection of videotaped neonatal seizures based on motion segmentation methods. Clin. Neurophys. **117**(7), 1585–1594 (2006)
16. Kusmakar, S., Karmakar, C.K., Yan, B., O'Brien, T.J., Muthuganapathy, R., Palaniswami, M.: Automated detection of convulsive seizures using a wearable accelerometer device. IEEE Trans. Biomed. Eng. **66**(2), 421–432 (2018)
17. Moshé, S.L., Perucca, E., Ryvlin, P., Tomson, T.: Epilepsy: new advances. Lancet **385**(9971), 884–898 (2015)
18. Nashef, L., So, E.L., Ryvlin, P., Tomson, T.: Unifying the definitions of sudden unexpected death in epilepsy. Epilepsia **53**(2), 227–233 (2012)
19. Osman, N., Camporese, G., Coscia, P., Ballan, L.: SlowFast rolling-unrolling LSTMs for action anticipation in egocentric videos. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3437–3445 (2021)
20. Pérez, J.S., Meinhardt-Llopis, E., Facciolo, G.: Tv-l1 optical flow estimation. Image Process. On Line **2013**, 137–150 (2013)
21. Pérez-García, F., Scott, C., Sparks, R., Diehl, B., Ourselin, S.: Transfer learning of deep spatiotemporal networks to model arbitrarily long videos of seizures. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12905, pp. 334–344. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87240-3_32
22. Rashid, M., Singh, H., Goyal, V.: The use of machine learning and deep learning algorithms in functional magnetic resonance imaging-a systematic review. Expert Syst. **37**(6), e12644 (2020)
23. Shih, J.J., et al.: Indications and methodology for video-electroencephalographic studies in the epilepsy monitoring unit. Epilepsia **59**(1), 27–36 (2018)
24. Siddiqui, M.K., Islam, M.Z., Kabir, M.A.: A novel quick seizure detection and localization through brain data mining on ECoG dataset. Neural Comput. Appl. **31**, 5595–5608 (2019)
25. Sivathamboo, S., et al.: Cardiorespiratory and autonomic function in epileptic seizures: a video-EEG monitoring study. Epilepsy Behav. **111**, 107271 (2020)
26. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems. vol. 30 (2017)
27. Wang, L., et al.: Temporal segment networks for action recognition in videos. IEEE Trans. Pattern Anal. Mach. Intell. **41**(11), 2740–2755 (2018)
28. Wang, X., et al.: OadTR: online action detection with transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 7565–7575 (2021)
29. Yang, Y., Sarkis, R.A., El Atrache, R., Loddenkemper, T., Meisel, C.: Video-based detection of generalized tonic-clonic seizures using deep learning. IEEE J. Biomed. Health Inf. **25**(8), 2997–3008 (2021)
30. Yuan, Y., Xun, G., Jia, K., Zhang, A.: A multi-context learning approach for EEG epileptic seizure detection. BMC syst. Biol. **12**(6), 47–57 (2018)
31. Zheng, N., Song, X., Su, T., Liu, W., Yan, Y., Nie, L.: Egocentric early action prediction via adversarial knowledge distillation. ACM Trans. Multimedia Comput. Commun. Appl. **19**(2), 1–21 (2023)