# Epileptic seizure detection with Convolutional Neural Networks and the Continuous Wavelet Transform

Carlos E R. Cardoso*, Aldo A. Díaz, Aline Pansani, Emerson Ntikawa, Diego Colugnati, Pedro Braga, Cláudio Quintino
Universidade Federal de Goiás (UFG)
Marcos Aureliano
Universidade Estadual Paulista (UNESP)

***Abstract*—The study of epileptic seizure often involves animal models to simulate the human behavior. Such models demand monitoring the evolution of the animal behavior continuously. Detecting seizure in this setup remains a challenge, because it typically requires trained personnel to annotate video sequences looking for the timestamps of seizure events. Deep Learning methods can help to solve this task in a more automatic and efficient manner due to their capacity of retrieving patterns from data. In this work, we conducted a pilot study to detect epileptic seizure from the images of small rodents using Convolutional Neural Networks (CNN) and the Continuous Wavelet Transform (CWT). We used the Social LEAP Estimates Animal Poses (SLEAP) framework for animal recognition to extract the morphological skeleton. Then, our CWT-CNN method used information of the frequency, magnitude and temporal evolution of head and thorax displacements to classify the animal behavior. The results showed a mean accuracy of 82.7% in the classification of epileptic seizure events.**

## I. Introduction

Epilepsy is a neurological disease with one of the highest incidences in the world [1]. Thus, it is essential to study the mechanisms behind the epilepsy genesis, the seizure generation, and the resistance against anti-epilepsy drugs. Animal models are the baseline in these studies. The pilocarpine-induced model [2] is one of the most famous models because it manifests spontaneous and random seizures, which are typical human conditions that can be classified respecting the 5-stage Racine scale [3]. In particular, the Racine stage 5 is characterized by a generalized tonic clonic seizure that presents a forelimb clonus, dorsal extension and rearing followed by loss of motor control and falling [4].

However, the randomness of the seizure events in the pilocarpine model difficult the prediction of its occurrence. The animals must be continuously monitored through video recordings, which requires trained personnel to manually annotate the occurrence of seizure events.

In such context, Deep Learning methods are helpful due to their high performance in pattern recognition tasks [5]. Since the pilocarpine model manifests abrupt motion in the thoracic region during seizures, the Deep Learning approach can be used to address the seizure detection problem.

In this pilot study, we used a Convolutional Neural Network (CNN) in combination with the Continuous Wavelet Transform (CWT) to automatically analyze the video recordings of rats with chronic epilepsy induced by pilocarpine. Our methodology involves methods for automatic animal recognition, posture tracking, extraction of motion signals, training and testing the CNN to classify animal behavior in Pilocarpine-induced rodents regarding epileptic seizures.

## II. Materials and methods

### A. Data set

The videos were recorded with resolution of 1280 x 720 pixels in the .avi extension. Five Wistar rats were video-monitored during eight months. The videos were captured with an RGB Camera Infra 1010 D Multi HD VHD. A total of 201 hours were recorded.

The scenes show five boxes containing the species, as depicted in Fig. 1. In order to analyze the individuals, we marked five regions of interest (ROI) that matched the boundaries of each box. The ROIs are separated and treated as independent videos.

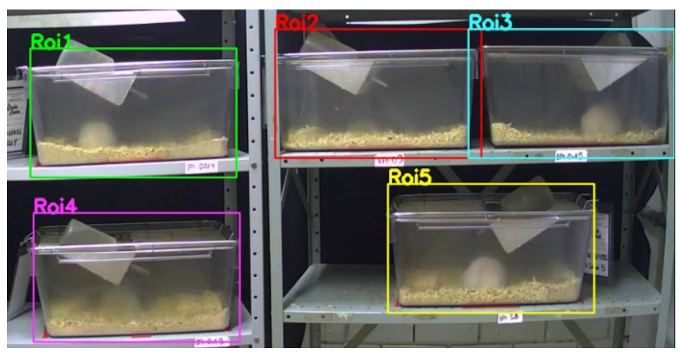


Figure 1. Location of the five ROIs.

### B. SLEAP – Estimation of the animal posture

The description of the animal posture can be related to several behaviors [6], [7]. Such description can be achieved with an Animal Posture Estimative (APE), in which the animal motion is tracked using specific points on the body that can be related with behavior information [8], [9]. Therefore, precise posture tracking is essential to build accurate models of the animal behavior.

One of the most advanced APE methods based on Deep Learning is the Convolutional Neural Network (CNN) [10] [11]. In this method, the keypoints are predicted using a statistical model that, once trained, it can infer keypoints on new datasets.

We used the Social LEAP Estimates Animal Poses (SLEAP), which is an open-source framework for posture

* Corresponding author e-mail: carlosribeiro.aca@gmail.com

classification [11]. The SLEAP framework allows posture estimation of multiple targets, even in the presence of complex social interactions. SLEAP is a pre-trained neural network and very user-friendly.

The first step was to manually label local keypoints on the animals. Since we used low-resolution images, we defined just two keypoints, the head and the thorax (see Fig. 2).

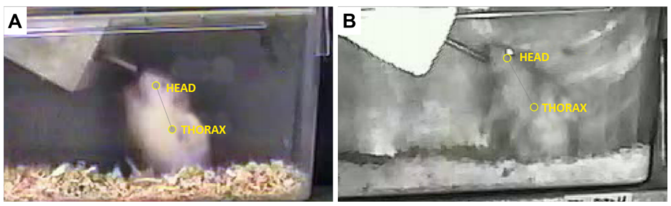


Figure 2. Local keypoints on two different animal postures.

Secondly, we set UNet as the CNN model in SLEAP (Fig. 3). This is a encoder-decoder network, originally created for image segmentation, that can achieve great precision in small datasets [12]. The UNet network can be trained in a Google Collaboratory environment, since SLEAP allows exporting the training parameters containing the network weights.

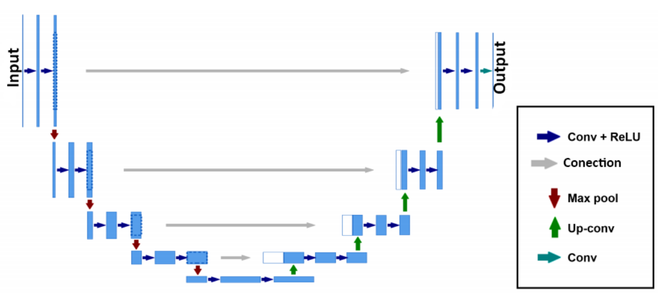


Figure 3. UNet code-encoder architeture example.

## *C. The CWT–CNN method*

The Continuous Wavelet Transform (CWT) is a signal analysis technique that transforms a temporal signal into a *scalogram* 4. The scalogram is an image containing both the temporal and frequency information of the signal. As an image, the CWT can be analyzed using a CNN to extract patterns from it.

However, instead of analyzing posture by means of $\boldsymbol{p}(t) = (X(t), Y(t))^T$, the 2D position of keypoints over time, we used

$$\boldsymbol{v}(t) = \frac{\mathrm{d}\boldsymbol{p}(t)}{\mathrm{d}t} = \left(\frac{\mathrm{d}X(t)}{\mathrm{d}t}, \frac{\mathrm{d}Y(t)}{\mathrm{d}t}\right)^T = \begin{pmatrix} v_X(t) \\ v_Y(t) \end{pmatrix}, \quad (1)$$

the velocity components of motion for each frame. Our analysis is based on a key observation: Epileptic seizure implies abrupt changes of position, thus, the velocity can better highlight such changes.

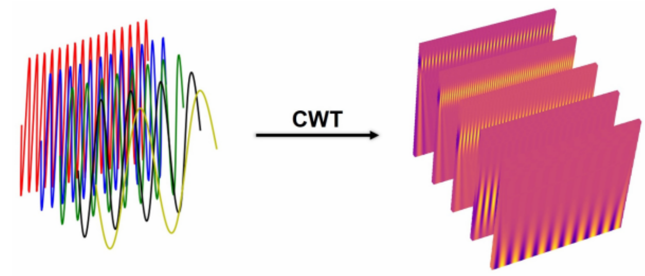


Figure 4. Scalogram outputed by CWT method from temporal signals.

Then, we applied a mexican and a morlet wavelet kernel [13], generating $N$ image matrices of $60 \times 240$ size. The first dimension is attributed to the wavelet scale. A small size was chosen to better visualize the low frequency components. The second dimension refers to the temporal duration of the video sequence. We used $40$ seconds with a total of $6$ frames per sequence. In order to keep the computational cost low, we downsampled the second dimension in a factor of $2$. The resulting signal is a structure of $N$ channels of $60 \times 120$ size.

The last layer of the CNN was designed to classify patterns in 6 possible animal behavior:

1) Class 1 – Epileptic seizure
2) Class 2 – Water ingestion
3) Class 3 – Food ingestion
4) Class 4 – Exploration
5) Class 5 – Grooming
6) Class 6 – No motion

The CNN architecture was composed of 3 convolutional layers using the Rectified Linear Unit (ReLU) [14] as activation function with a Batch Normalization on each convolutional layer. Data is vectorized to generate a $1D$-vector. Then, 3 processing layers are applied, 2 with the ReLU and 1 with SoftMax activation functions.

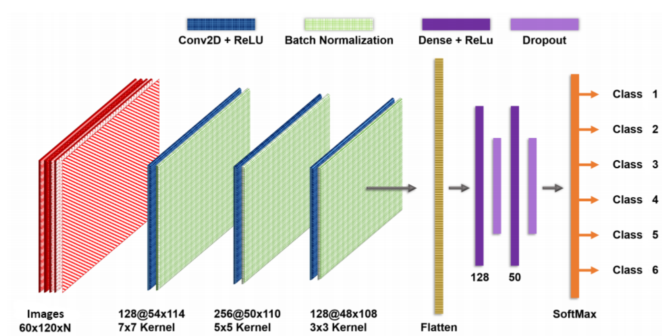


Figure 5. The CWT–CNN architecture.

The CWT–CNN (Fig. 5) was trained using the Categorical Cross Entropy as loss function and the Adaptive Moment Estimation as as the optimizer to control the update of the network weights during the learning process.

## III. RESULTS

### A. SLEAP

We used active training in the APE model, that is, we prior informed to the neural network the real head and thorax positions. Firstly, 292 frames were used to learn the tracking of keypoints in the animal's body Another 163 new frames were used as the testing set. The training was repeated six times, varying the frames exposed to the neural network. At the end 2203 frames were utilized.

The results of the keypoint detection process are shown in Fig. 6. In order to assess the APE capability of detecting the head and thorax correctly, we exposed the network with 200 new frames and applied a simple metric:

$$\text{Correct detections} = \left(1 - \frac{N_{\text{errors}}}{200}\right) * 100\,, \qquad (2)$$

where $N_{\text{errors}}$ is the number of incorrect keypoint detections (false positives and false negatives).

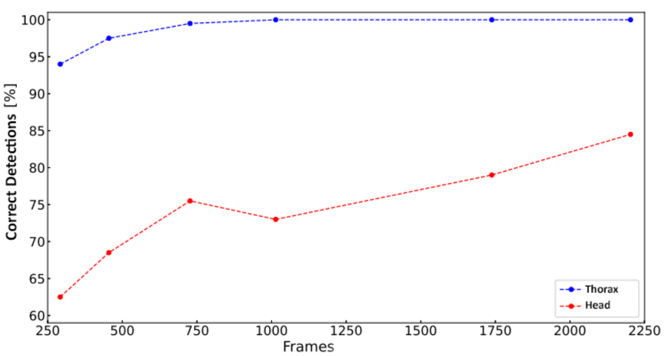


Figure 6. Total number of correct detections per frame.

As an error metric, we also computed the Euclidean distance between the actual location of the keypoints and the predicted one (Fig. 7). The actual position was previously marked to create a source of comparison.

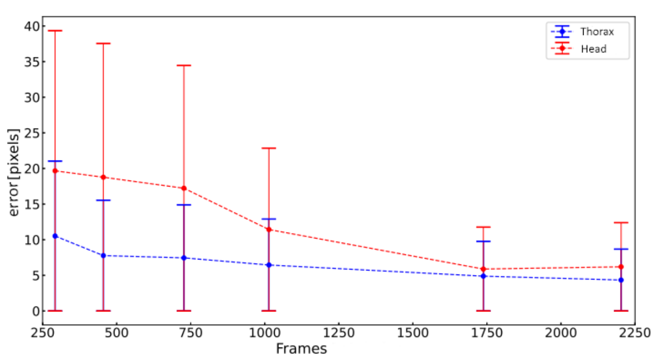


Figure 7. Difference between real local points and predicted ones.

At a maximum frame exposure (frame 2203), we obtained an keypoint tracking error of $4.3 \pm 3.5$ for the thorax and $19.7 \pm 24.9$ for the head. The error distribution, however, shows that 75% of the errors are within 7 pixels of distance at maximum as shown in (Fig.8). Since the images we used are of low-resolution and, given that the SLEAP framework can reach a 3-pixel-distance error in high resolution images, we concluded that the model demonstrated satisfactory results given the study limitations.

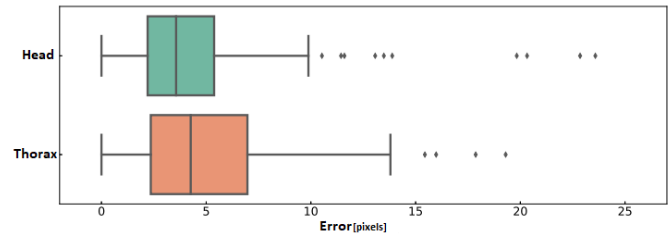


Figure 8. Error margin from the best estimative model.

### B. CWT–CNN

A new data set consisting of 232 frames was considered in order to evaluate the performance of our model. The data set included the following behaviors (manually classified): 52 epileptic seizure; 30 water ingestion; 28 food ingestion; 42 exploration; 33 grooming; 47 no motion.

A Savitzky-Golay filter [15] from the SciPy library [16] was applied to each frame in order to extract the 4 motion signals: The $X$ and $Y$ velocities of the thorax and the head. With these information, we designed 4 different weighted combinations and used them as input signals to the CWT. The signals are summarized in Table I. The $H$ subscript refers to the head keypoints and the $T$ subscript refers to the thorax keypoints.

Table I
THE FOUR MOTION SIGNALS AND THEIR COMBINATIONS.

| Combi1 | Combi2 | Combi3 | Combi4 |
|---|---|---|---|
| $\frac{dX_T}{dt}$ | $\frac{dX_T}{dt}$ | $\frac{dX_T}{dt}$ | $\frac{dX_T}{dt}$ |
| $\frac{dY_T}{dt}$ | $\frac{dY_T}{dt}$ | $\frac{dY_T}{dt}$ | $\frac{dY_T}{dt}$ |
| $\frac{dX_H}{dt}$ | $\frac{dX_T}{dt} \cdot \frac{dY_T}{dt}$ | $\frac{dX_T}{dt} \cdot \frac{dY_T}{dt}$ | $\frac{dX_T}{dt} \cdot \frac{dY_T}{dt}$ |
| $\frac{dY_H}{dt}$ | $\frac{dY_H}{dX_H}$ | $\frac{dY_H}{dX_H}$ | $\left(\frac{1}{4}\frac{dX_T}{dt}\right)^2 + \left(\frac{dY_T}{dt}\right)^2$ |
| $\ldots$ | $\ldots$ | $\frac{dY_H}{dt}$ | $\frac{dY_H}{dt}$ |

Since the *dorsal extension* is a key factor in determining epileptic seizure, the thorax and head velocities in the $Y$-axis were the most informative signals utilized in the combinations. However, we preserved the entire information to provide further the model with further detail as a robustness heuristic that may increase the chance of correctly classifying seizure events.

We applied $k$-fold cross validation during training. The data set was divided into $k = 5$ folds of same size. Training was conducted separately on each fold using $k-1$ of $k$ folds each time. The last fold was used for testing.

We may also consider the randomness associated with the training of neural networks. On each run, the training process initializes with different network weights and, despite using the same data set, different results can be achieved. Due to this, we run the experiment for 10 iterations on the same data set and we took the average of the results. We used the accuracy as performance parameter, *i.e.*, the rate of correct predictions against actual behavior for each class (Table II).

Table II
ACCURACY LEVEL PER FOLD OF EACH COMBINATION.

| **Fold** | **Combi1 (%)** | **Combi2 (%)** | **Combi3 (%)** | **Combi4 (%)** |
|---|---|---|---|---|
| 1 | 59.2 | 58.1 | 57.7 | 65.8 |
| 2 | 74.9 | 73.6 | 68.3 | 80.0 |
| 3 | 87.2 | 83.9 | 81.1 | 88.3 |
| 4 | 81.5 | 86.1 | 76.8 | 92.4 |
| 5 | 86.3 | 85.5 | 76.3 | 87.2 |
| Mean | 77.8 | 77.4 | 72.0 | 82.7 |

## IV. DISCUSSION

We designed a posture model in order to detect epileptic seizures in small rodents. We tracked the position of four keypoints on the animal body, the $X$ and $Y$ positions from the thorax and the head. Our model was capable of fully tracking the thorax keypoints and partially tracking the head keypoints ($\approx 84.5\%$) in all the $292$ frames. The detection errors resulted in $4.3 \pm 3.5$ of pixel distance for the thorax keypoints and $6.2 \pm 10.2$ for the head keypoints. These results show a promising room for improvement, considering that our images were captured with low resolution and the SLEAP framework may improve its tracking performance as the image resolution increases.

Regarding our method for the classification of seizure events, we designed a model based on a combination of specific motion signals and used them as inputs to convolutional neural network to classift $5$ different types of animal behaviou. To test the model, we used a data set with $52$ epileptic seizure events. The best models were Combination 1 and Combination 4, which achieved a total mean accuracy of $77.8\%$ and $82.7\%$, respectively.

## V. CONCLUSIONS

Although we used a small data set, our CWT–CNN method was proven successful in classifying epileptic seizure in animal behavior. Potential strategies for improvement in future include collecting a larger data set to increase the model accuracy.

A detailed analysis reveals some issues regarding the individual performance of each signal combination. For instance, Combination 1 detected all the epileptic seizures without false positives. Combination 4 also detected all the epileptic seizures, but with the presence of false positives. Despite this, both combinations can be accepted because they fully detected seizure events without yielding any false negative. Therefore, we assure that the model did not bias our quantitative study and, any false positive produced can later be treated by human analysis.

In order to achieve a fully automatic model, further improvements need to be investigated. In particular, using a larger data set to provide robustness against image noise and different scene configurations.

## REFERENCES


[1] A. O. T. J. J. N. S. I. E. d. C. M. P. P. Devinsky, Orrin; Vezzani, "Epilepsy," *Nature Reviews Disease Primers*, vol. 4, p. 18024, 2018.

[2] W. A. T. E. A. C. M. S. S. J. C. Z. K. L. Turski, "Limbic seizures produced by pilocarpine in rats: Behavioural, electroencephalographic and neuropathological study," *Behavioural Brain Research*, vol. 9, pp. 315–335, 1983.

[3] R. J. Racine, "Modification of seizure activity by electrical stimulation: Ii. motor seizure," *Electroencephalography and Clinical Neurophysiology*, vol. 32, pp. 0–294, 1972.

[4] G. Curia, D. Longo, G. Biagini, R. S. Jones, and M. Avoli, "The pilocarpine model of temporal lobe epilepsy," *Journal of Neuroscience Methods*, vol. 172, no. 2, pp. 143–157, 2008.

[5] V. C. B. K. D. N. S. K. A. F. D. E.-G. S. W. S. Antonopoulos, Ioannis; Robu, "Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review," *Renewable and Sustainable Energy Reviews*, vol. 130, p. 109899, 2020.

[6] J. W. M. M. Pereira, Talmo D.; Shaevitz, "Quantifying behavior to understand the brain," *Nature Neuroscience*, 2020.

[7] U. Klibaite and J. W. Shaevitz, "Paired fruit flies synchronize behavior: Uncovering social interactions in drosophila melanogaster," *PLOS Computational Biology*, vol. 16, pp. 1–21, 10 2020.

[8] D. E. W. L. K. M. W. S. S.-H. M.-M. S. J. W. Pereira, Talmo D.; Aldarondo, "Fast animal pose estimation using deep neural networks," *Nature Chemical Biology*, vol. 16, pp. 117–125, 2019.

[9] S. Günel, H. Rhodin, D. Morales, J. Campagnolo, P. Ramdya, and P. Fua, "Deepfly3d, a deep learning-based approach for 3d limb and appendage tracking in tethered, adult *Drosophila*," *eLife*, vol. 8, p. e48571, oct 2019.

[10] P. C. K. M. A. T. M. V. N. M.-M. W. B. M. Mathis, Alexander; Mamidanna, "Deeplabcut: markerless pose estimation of user-defined body parts with deep learning," *Nature Neuroscience*, 2018.

[11] T. D. Pereira, N. Tabris, A. Matsliah, D. M. Turner, J. Li, S. Ravindranath, E. S. Papadoyannis, E. Normand, D. S. Deutsch, Z. Y. Wang, G. C. McKenzie-Smith, C. C. Mitelut, M. D. Castro, J. D'Uva, M. Kislin, D. H. Sanes, S. D. Kocher, S. S-H, A. L. Falkner, J. W. Shaevitz, and M. Murthy, "Sleap: A deep learning system for multi-animal pose tracking," *Nature Methods*, vol. 19, no. 4, 2022.

[12] J. M. Graving, D. Chae, H. Naik, L. Li, B. Koger, B. R. Costelloe, and I. D. Couzin, "Deepposekit, a software toolkit for fast and robust animal pose estimation using deep learning," *eLife*, vol. 8, p. e47994, oct 2019.

[13] X. M. H. R. Z. O. W. W. K. Ma, "The use of the mexican hat and the morlet wavelets for detection of ecological patterns," *Plant Ecology*, vol. 179, pp. 1–19, 2005.

[14] A. F. Agarap, "Deep learning using rectified linear units (relu)," *arXiv preprint arXiv:1803.08375*, 2018.

[15] R. W. Schafer *et al.*, "What is a savitzky-golay filter," *IEEE Signal processing magazine*, vol. 28, no. 4, pp. 111–117, 2011.

[16] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.