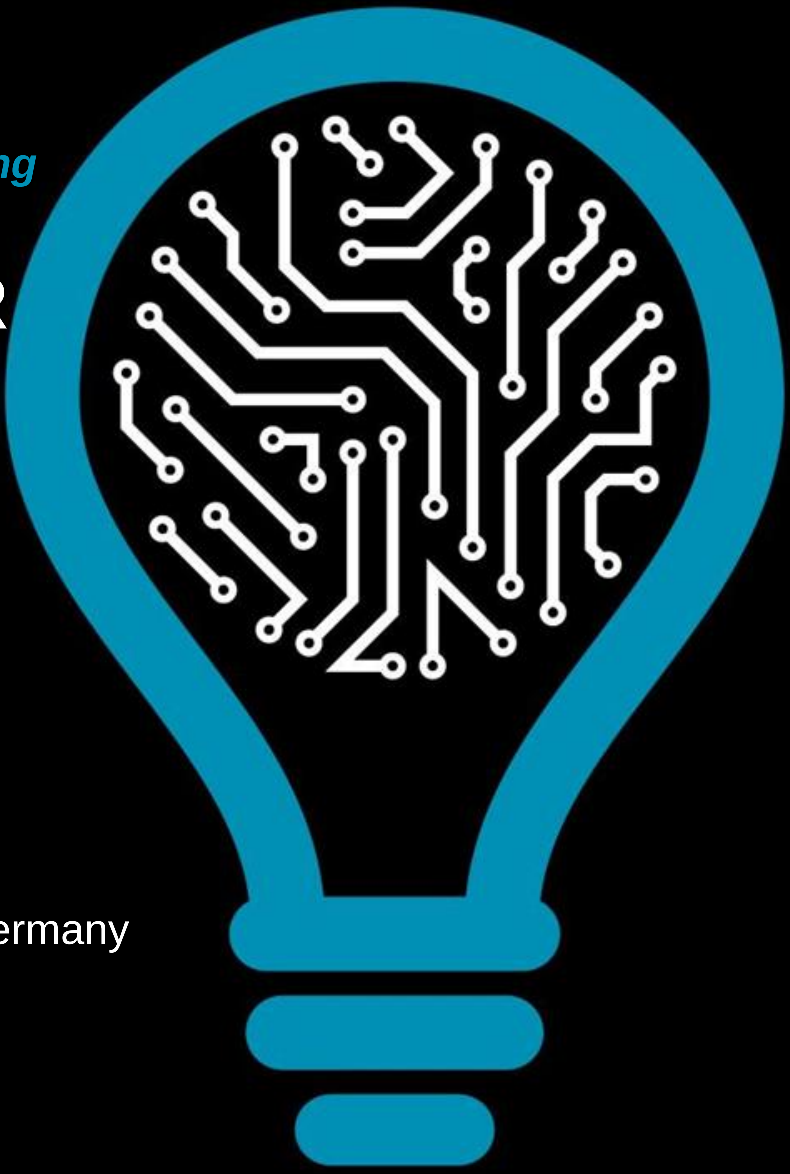
ICASSP 2020 Barcelona

45th International Conference on Acoustics, Speech, and Signal Processing
May 4-8 2020

A DEEP LEARNING ARCHITECTURE FOR EPILEPTIC SEIZURE CLASSIFICATION BASED ON OBJECT AND ACTION RECOGNITION



Tamás Karácsony[1], Anna Mira Loesch-Biffar[2], Christian Vollmar[2], Soheyl Noachtar[2], João Paulo Silva Cunha[1,3]

[1] Center for Biomedical Engineering Research, INESC TEC, Porto, Portugal
[2] Epilepsy Center, Department of Neurology, University of Munich, Munich, Germany
[3] Faculty of Engineering (FEUP), University of Porto, Porto, Portugal

INESCTEC INSTITUTE FOR SYSTEMS AND COMPUTER ENGINEERING, TECHNOLOGY AND SCIENCE

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

LMU KLINIKUM DER UNIVERSITÄT MÜNCHEN

FCT Fundação para a Ciência e a Tecnologia MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

UNIÃO EUROPEIA Fundo Europeu de Desenvolvimento Regional

Introduction-Epilepsy

Motivation & Objectives

State of the art

Methods

Results

Conclusion

Future work

Introduction – Epilepsy

- Neurological disorder that affects 0.5-1% of the world population
- Epilepsy monitoring units
  - Rely on visual inspection
  - 2Dvideo-EEG data for diagnosis
  - Patient’s movements of interest (MOIs)
- Subjective method
- Requires a lot of resources

Motivation and Objectives

- Recent improvements in machine learning
  - Human action recognition
  - Computer vision
- Support diagnosis with machine learning
- Need for automatic epilepsy classification

State of the art

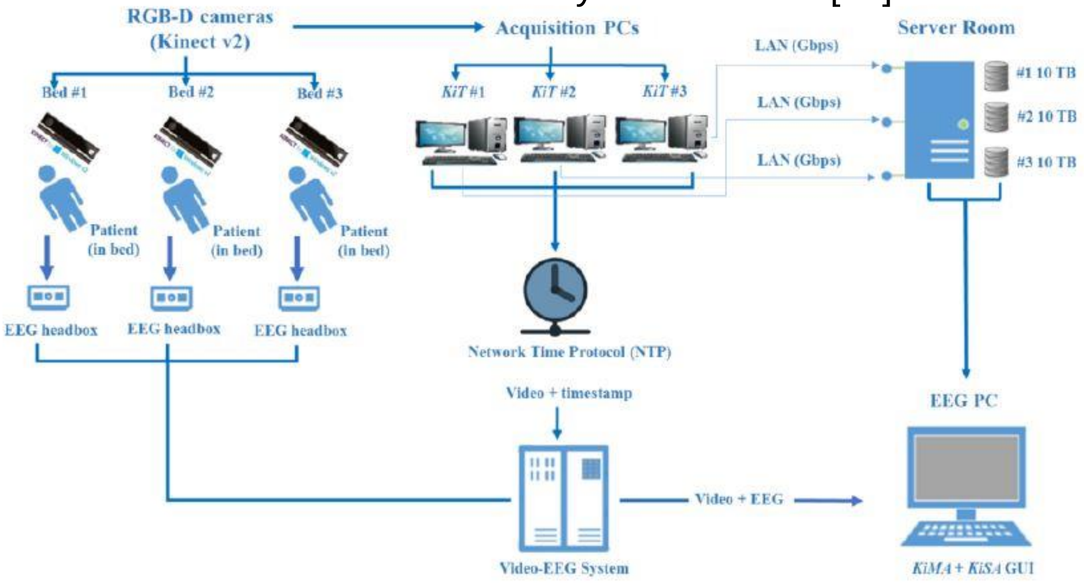
Author	Classes	Performance	Notes
Achilles et al. [17, 18]	Seizure No seizure	AUC: 0.78	Single frame approach (posture recognition)
Ahmedt-Aristizabal et al. [19]	MTLE ETLE	Average accuracy: 53.39%-56.31%	Face body and hand inputs, very high std
Maia et al. [21]	TLE ETLE	AUC 0.65	Probably overfits
This work	**TLE** **FLE**	**f1-score: 0.844±0.042** **AUC: 0.90±0.04**	-

- Hierachical model proposed by Ahmedt-Aristizabal et. Al [19]
  - Detection and tracking algorithms (patient, face, limbs, head, hand)
  - Convolutional NN, Recurrent NN (LSTM), 2D video
  - Limited success
- CNN based method Achilles et al. [17]
  - Depth + IR videos
  - Limited, insufficent information
- Maia et al. [21]
  - Inception-V3 object recognition feature extraction on IR videos
  - Author suggest overfit due to class imbalance

Methods – Datasets-I

- 3D-video (RGB-D) Neurokinect 3.0 dataset
  - Frontal Lobe Epilepsies (FLE), Temporal Lobe Epilepsies (TLE)
  - Infrared (IR) videos
  - 126 seizures from 35 patients

Neurokinect 3.0 by Cunha et. Al. [13]



Example video from the dataset (up) and main metrics (down)

Class name	Frontal Lobe Epilepsy (FLE)	Temporal Lobe Epilepsy (TLE)
Included seizures	FLE, right FLE, left FLE	TLE, right TLE, left TLE
Number of patients	20	15
Number of seizures	85	41
Total clinical length [s]	2587	3116
Average clinical length [s]	30.4	76.0
Minimal clinical length [s]	1.4	6.3
Maximal clinical length [s]	187.9	225.9
Resolution	512x424 16bit	
Sampling frequency	30 fps	

Methods – Datasets-II

- Datasets utilized for transfer learning
  - MS-COCO [25]
    - Image segmentation
    - static
  - ImageNet [14]
    - Largest image dataset
    - static
  - Kinetics [15]
    - Human actions (400 class)
    - dynamic

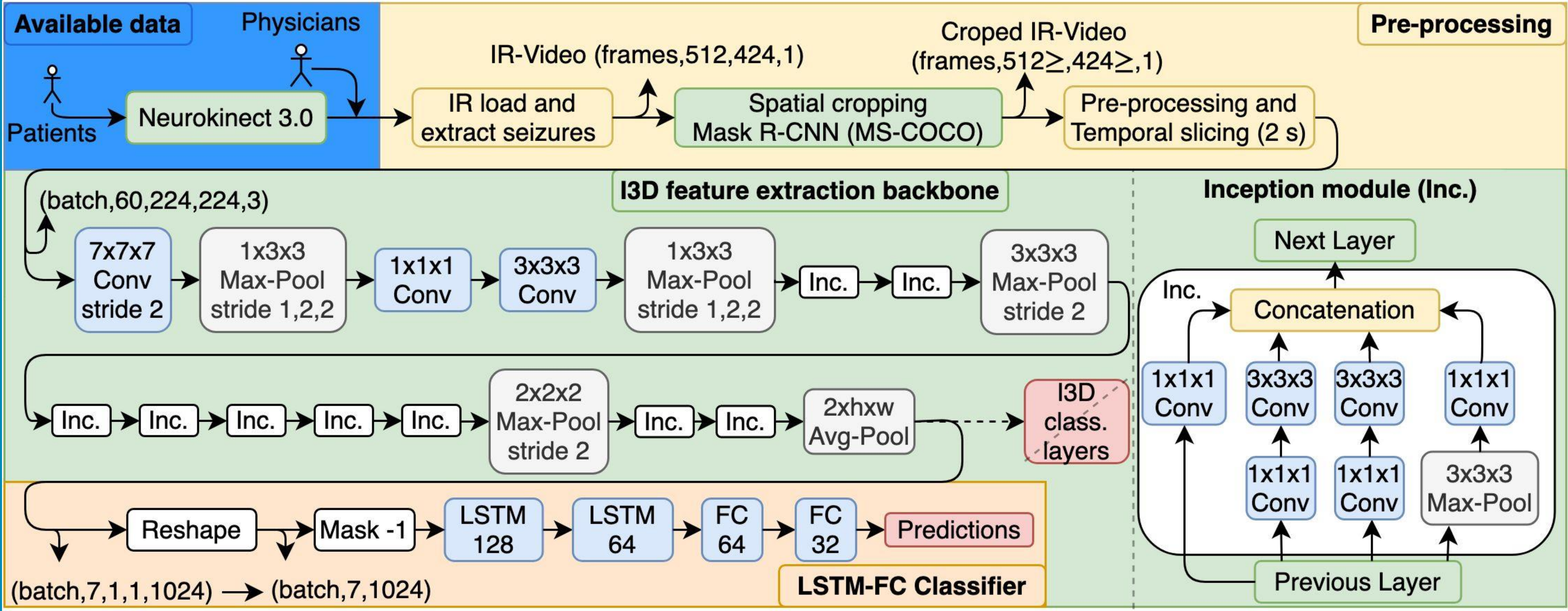
MS-COCO example image #449661 [25]

Example videos from Kinetics [15]

Methods – Deep learning

- Architectures
  - Mask R-CNN based algorithm for image segmentation (bed+person) [22]
    - MS-COCO pre-training
  - Inflated Inception-V1 architecture (I3D) for feature extraction [16]
    - ImageNet + Kinetics pre-training
    - Apparent (spatial) information
    - Motion information
    - Short-term temporal information
  - LSTM classifier
    - Long term temporal connections

The full architecture



LSTM Classifier

- LSTM layers: 128 & 64 units
  - Recurrent dropout (0.3)
- Fully connected layers: 64 & 32 units
  - He uniform initializer
  - ReLU activation
- Regularization:
  - Batch normalizition
  - Dropout (0.5)
  - L2 regularization

PARAMTERES OF THE DEVELOPED LSTM FEATURE CLASSIFIER

Layer (type)	**Output Shape**	**Param #**
Mask (Masking)	(None, 7, 1024)	0
BN_1 (BatchNormalization)	(None, 7, 1024)	4096
DO_1 (Dropout)	(None, 7, 1024)	0
LSTM_1 (LSTM)	(None, 7, 128)	590336
BN_2 (BatchNormalization)	(None, 7, 128)	512
DO_2 (Dropout)	(None, 7, 128)	0
LSTM_2 (LSTM)	(None, 64)	49408
BN_3 (BatchNormalization)	(None, 64)	256
DO_3 (Dropout)	(None, 64)	0
FC_1 (Dense)	(None, 64)	4160
BN_4 (BatchNormalization)	(None, 64)	256
DO_4 (Dropout)	(None, 64)	0
FC_2 (Dense)	(None, 32)	2080
Class_out (Dense)	(None, 1)	33
Total params: 651,137		
Trainable params: 648,577		
Non-trainable params: 2,560		

INESCTEC

Training & Evaluation

- Temporal slicing (2 [s] segments)
  - Counteract class imbalance
  - Data augmentation
- Weighted binary cross-entropy
- Adam optimizer
- 2000 epochs (max, early stoping)
- Batch size: 500 samples

- Mask R-CNN visual confirmation
- 5-fold cross validation
(Mask R-CNN+ I3D+LSTM)
  - F1 score
  - Precision
  - Recall

Results – Mask R-CNN

- Mask R-CNN visual confirmation
  - 122 of 126 correct (96.83 %)
- Under or miscrop
  - Only due to heavy occlusions

(a) Before crop

(b) After crop

(a) Before (under)crop

(b) After (under)crop

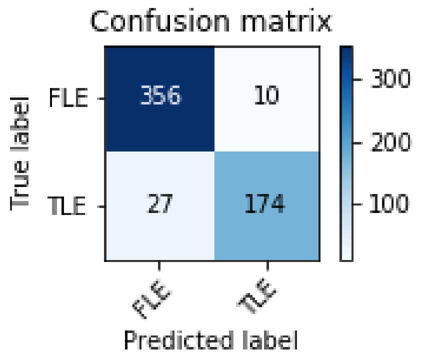
Results - Classification of I3D features

- 5-fold cross-validation average of macro average metrics
  - f1-score: 0.844 ± 0.042
  - Precision: 0.857 ± 0.042
  - Recall: 0.838 ± 0.041

EXAMPLE METRICS OF THE BEST FOLD IN THE 5-FOLD CROSS VALIDATION

	f1-score	precision	recall	support
FLE	0.930	0.973	0.951	366
TLE	0.946	0.866	0.904	201
macro avg	0.938	0.919	0.927	567
weighted avg	0.935	0.935	0.934	567
accuracy	0.935			567

Author	Classes	Performance	Notes
Achilles et al. [17, 18]	Seizure No seizure	AUC: 0.78	Single frame approach (posture recognition)
Ahmedt-Aristizabal et al. [19]	MTLE ETLE	Average accuracy: 53.39%-56.31%	Face body and hand inputs, very high std
Maia et al. [21]	TLE ETLE	AUC 0.65	Probably overfits
This work	**TLE** **FLE**	**f1-score: 0.844±0.042** **AUC: 0.90±0.04**	-



Conclusion

- An end-to-end deep learning approach was proposed (Mask R-CNN + I3D + LSTM)
  - Motion based binary classification
  - Frontal and Temporal Lobe Epilepsies
- Promising classification results
- This contact-less sensor (IR) based classification has the potential to support physicians with diagnostic decisions and might be applied for online applications in epilepsy monitoring units.

Future work

- Additional data streams
  - Depth data
- Improved preprocessing and data augmentation
- Extending dataset
- Adding non-seizure class

INESCTEC

Acknowledgements

- This project was partially financed by the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, through national funds, and co-funded by the FEDER, where applicable, and by National Funds through the Portuguese funding agency, FCT—Fundação para a Ciência e a Tecnologia within project POCI-01-0145-FEDER-028618 (PTDC/CCI-COM/28618/2017).

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

UNIÃO EUROPEIA
Fundo Europeu
de Desenvolvimento Regional

INESCTEC

References

- [13] Hugo Miguel Pereira Choupina, Ana Patrícia Rocha, José Maria Fernandes, Christian Vollmar, Soheyl Noachtar, and João Paulo Silva Cunha. Neurokinect 3.0: Multi-bed 3dvideo-eeg system for epilepsy clinical motion monitoring. Studies in health technology and informatics, 247:46–50, 2018.
- [14] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in 2009 IEEE Conference on Computer Vision and Pattern Recognition. jun 2009, IEEE.
- [15] W. Kay, J. Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijayanarasimhan, F. Viola, T. Green, T. Back, P. Natsev, M. Suleyman, and A. Zisserman, “The Kinetics Human Action Video Dataset,” arXiv, 2017.
- [16] J. Carreira and A. Zisserman, “Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). jul 2017, IEEE.
- [17] F. Achilles, F. Tombari, V. Belagiannis, A. M. Loesch, S. Noachtar, and N. Navab, “Convolutional neural networks for real-time epileptic seizure detection,” Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, vol. 6, no. 3, pp. 264–269, jul 2016.
- [18] F. Achilles, V. Belagiannis, F. Tombari, A. Loesch, J. Cunha, N. Navab, and S. Noachtar, “Deep convolutional neural networks for automatic identification of epileptic seizures in infrared and depth images,” Journal of the Neurological Sciences, vol. 357, pp. e436, oct 2015.
- [19] David Ahmedt-Aristizabal, Clinton Fookes, Simon Denman, Kien Nguyen, Tharindu Fernando, Sridha Sridharan, and Sasha Dionisio. A hierarchical multimodal system for motion analysis in patients with epilepsy. Epilepsy & Behavior, 87:46–58, oct 2018.
- [21] P. Maia, E. Hartl, C. Vollmar, S. Noachtar, and J. P. S. Cunha, “Epileptic seizure classification using the NeuroMov database,” in 2019 IEEE 6th Portuguese Meeting on Bioengineering (ENBENG). feb 2019, IEEE.
- [22] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask r-CNN,” in 2017 IEEE International Conference on Computer Vision (ICCV). oct 2017, IEEE.
- [25] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft COCO: Common Objects in Context,” in Computer Vision – ECCV 2014, pp. 740–755. Springer International Publishing, 2014.

Thank you for your attention!