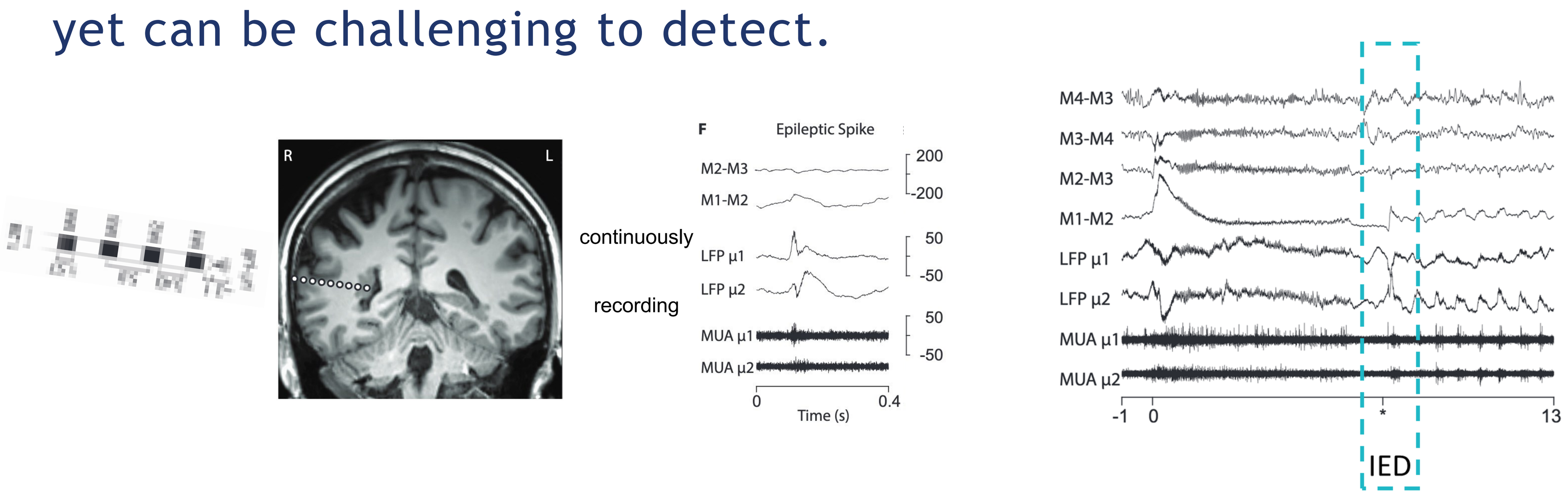
A human-in-the-loop Active-Deep Learning Framework for Automated Interictal Epileptiform Discharges (IED) Detection from EEG Data

Shen Liang[1], Qitong Wang[2], Delphine Roussel[3], Stephen Whitmarsh[4], Katia Lehongre[4], Alexandra Levchenko[5], Vincent Navarro[4], Themis Palpanas[1,6]

1. LIPADE, Université Paris Cité
2. Harvard University
3. Centre Institut de l'Audition – Institut Pasteur
4. Sorbonne Université, Institut du Cerveau - ICM, Inserm, CNRS, APHP, Hôpital de la Pitié Salpêtrière
5. Institut Supérieur d'Electronique de Paris – ISEP
6. Institut Universitaire de France (IUF)

Introduction

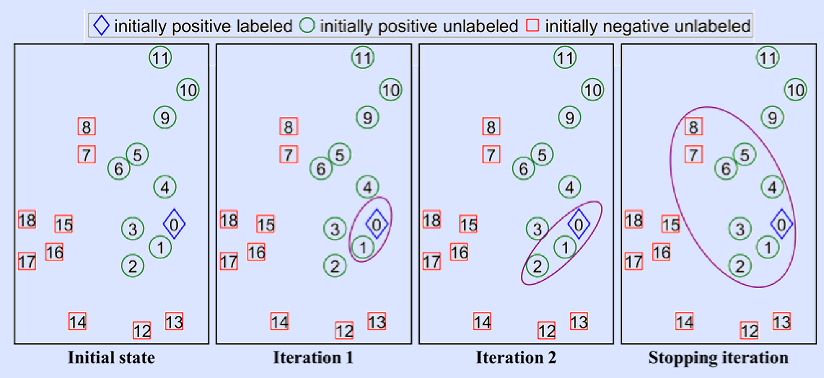
- Interictal Epileptiform Discharges (IEDs) in EEG data are critical biomarkers for diagnosis and treatment of epilepsy, yet can be challenging to detect.

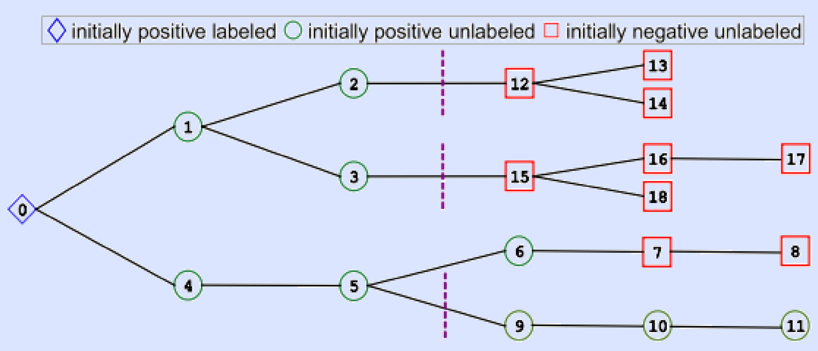


- Numerous deep learning (DL) methods for automated IED detection, often framing IED detection as a two-class classification problem between IED and non-IED EEG snippets.
- Existing DL methods have limited capacity in addressing the inherent imbalanced nature of IED detection.
  - IEDs are rare events while also exhibiting a certain level of heterogeneity.
  - Most existing DL methods require (potentially large numbers of) pre-existing, manually labeled training data, with limited consideration as to whether the labeled examples can best represent the IED and non-IED data spaces, especially given data imbalance.
  - Most existing DL methods have not been tailored to imbalance-aware evaluation metrics (e.g. the F-score) in their training processes.
- We propose a novel human-in-the-loop active-deep learning framework for automated IED detection that, by effectively addressing the aforementioned data imbalance issue, significantly reduce the number of manual labels required while obtaining high detection accuracy.

Methodology (continued)

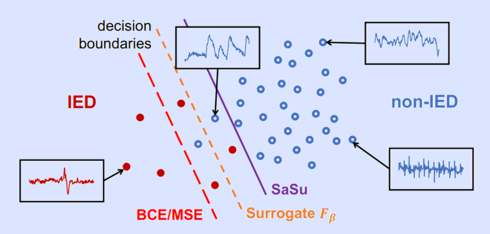
- When selecting the example for user to label manually, the proposed pipeline strives to ensure as many and as diverse IED examples are in the eventual training set.
  - Initially, the pipeline prioritizes querying high-amplitude examples which also have relatively small distances to their nearest neighbor in the dataset (which indicates their morphologies are relatively defined), drawing on the domain knowledge that IEDs typically have high amplitudes and spike-and-slow-wave morphologies. This way, we can maximize the number of IED examples that are manually labeled.
  - The pipeline then draws on a tree-based algorithm to infer the unlabeled examples, as the different tree branches naturally encode heterogeneity of IEDs.



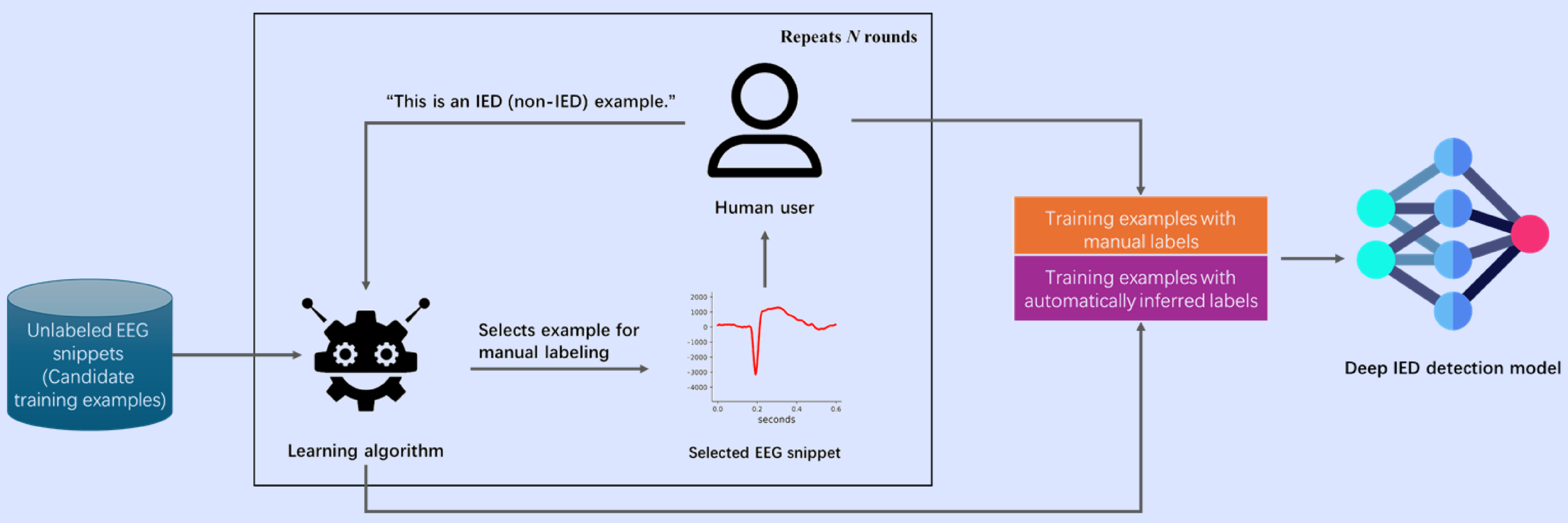


  - Further user interactions are conducted to refine the automatic inference results while filtering the examples with low-confidence inferred labels.
- The selected examples are used to train a convolutional DL model with a novel loss function that can directly optimize the F-score of IED detection.

$$L_{SaSu}(f(x;\Theta), y) = -\, y \log f(x;\Theta) + (1-y)\log\left(\beta^2 \frac{\bar{p}_{1*,s}}{\alpha(1-\bar{p}_{1*,s})} + f(x;\Theta)\right)$$



Methodology



- Starting from a pool of unlabeled EEG snippets as candidates for training data, the proposed pipeline interactively queries the user for the labels of a small proportion (e.g. 10%) of examples deemed most informative to train the DL model.
- Based on the user-labeled examples, the pipeline automatically infers the labels of the remaining examples. The manually labeled examples, along with part of the automatically inferred examples with high confidence, are used to train the eventual DL model.

Experimental Results

- Experiments carried out for 7 human subjects and 3 epileptic mice.
- Using F1-score as evaluation metric

	H1	H2	H3	H4	H5	H6	H7	MO 1	MO2	MO3
FS	0.96	0.89	0.90	0.84	0.91	0.80	0.85	0.90	0.70	0.93
SOTA	0.91	0.88	0.66	0.79	0.77	0.76	0.82	0.89	0.61	0.76
Ours	0.95	0.84	0.88	0.84	0.80	0.80	0.88	0.89	0.67	0.77

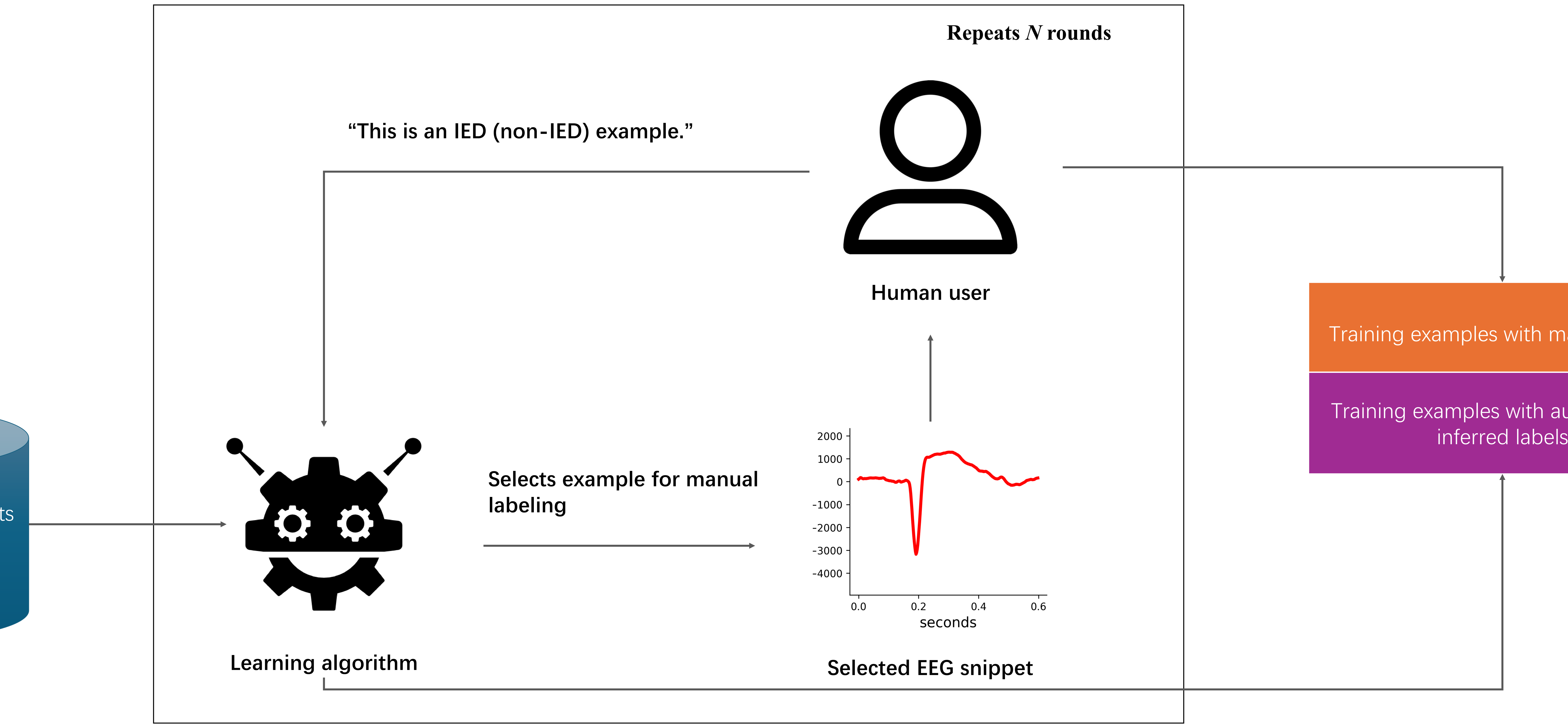
(FS = Manually labeling all candidate examples; SOTA = Best baseline method with 10% manual labeling; Ours = our pipeline with 10% manual labeling)

- By manually labeling just 10% of all candidate training examples, the proposed pipeline can obtain IED detection results comparable to those when all candidate examples are manually labeled.

	H1	H2	H3	H4	H5	H6	H7	MO 1	MO2	MO3
BCE	0.91	0.88	0.66	0.80	0.76	0.79	0.73	0.78	0.53	0.76
Ours	0.95	0.84	0.88	0.84	0.80	0.80	0.88	0.89	0.67	0.77

(BCE = Binary Cross Entropy loss; Ours = our F1-score-aware loss; both with 10% manual labeling)

- Compared with the conventional BCE loss, the proposed F-score-aware loss function can lead to better detection results.



	H1	H2	H3	H4	H5	H6	H7	MO 1	MO2	MO3
FS	0.96	0.89	0.90	0.84	0.91	0.80	0.85	0.90	0.70	0.93
SOTA	0.91	0.88	0.66	0.79	0.77	0.76	0.82	0.89	0.61	0.76
Ours	0.95	0.84	0.88	0.84	0.80	0.80	0.88	0.89	0.67	0.77

	H1	H2	H3	H4	H5	H6	H7	MO 1	MO2	MO3
BCE	0.91	0.88	0.66	0.80	0.76	0.79	0.73	0.78	0.53	0.76
Ours	0.95	0.84	0.88	0.84	0.80	0.80	0.88	0.89	0.67	0.77