**S2. REQUIRED INFORMATION**

Q 10

The Data Report should include an overview of the data files and their formats. Has this information been included?

**Reviewer 4** | 19 Sep 2018 | 16:24

**#1**

No
There should be a clearer description of the filetypes/formats, like channel-based annotation and term-based annotation, how are they stored precisely? Everything in EDF? Together with data? Maybe mention that earlier in result section to avoid any confusion.

**Comment:** We have updated the information related to transcription file-formats in the first paragraph of the “Results” section. Other detailed information about the standards being used are provided either in readme files or in a spreadsheet anyway.

 **S3. OTHER COMMENTS**

Q 14

Please add here any further comments on this manuscript.

 **Reviewer 4** | 19 Sep 2018 | 16:24

**#1**

The dataset is very useful and overall nicely described.
Is the code available for the keyword matching etc.? Could be nice.
I think it would be worth adding to the discussion potential biases in the sampling of the seizures due to the algorithms used for finding them. Especially as finding seizures with a deep-learning method might make such seizures easier to later detect again with similar deep-learning methods. Due to the small number of seizure files identified in that way, this is maybe not a major issue, but still worth mentioning. And for the keyword-based one at least a small discussion if that might systematically find some seizures but not others would have helped me as well.

**Comment:**

>> Is the code available for the keyword matching etc.?

This project was funded from commercialization funds so we are obligated to keep our codes proprietary.

>> I think it would be worth adding to the discussion potential biases in the sampling of the seizures due
>> to the algorithms used for finding them. Especially as finding seizures with a deep-learning method
>> might make such seizures easier to later detect again with similar deep-learning methods.

This is a valid concern. But our data triaging methods don’t simply collect files where deep-learning algorithms found seizures. We develop database on patient bases. If at least one seizure is identified by our recognition system, we extract all the sessions and their records for that patient for the annotation process. This can add some bias for the deep-learning systems but can’t be very influential.

In-between and within the sessions, the electrographic morphologies could vary drastically depending on various factors such as medications, age, type of epilepsy, gap between two recording sessions, etc… So, detecting seizures from few files wouldn’t necessarily make all those seizure patterns easier to detect for ML algorithms. In fact, our evaluation set is designed to be as diverse as possible in this regard even when it contains only 50 patients.

>> And for the keyword-based one at least a small discussion if that might systematically find some
>> seizures but not others would have helped me as well.

At this stage, we have not evaluated the performance of our keyword search algorithm on missing seizures of a specific type. This off-the-shelf technology was developed and evaluated on a small amount data which was not enough to make any specific conclusions on the detection of specific seizure types. But they do tend to fail on sentences containing pseudo-negations (i.e. “the record does not rule out the presence of an ictal activity”).