ISRCTN18025327 https://doi.org/10.1186/ISRCTN18025327

# AI for epilepsy classification

| Submission date | Recruitment status | [_] Prospectively registered |
|---|---|---|
| 02/01/2024 | No longer recruiting | [_] Protocol |
| **Registration date** | **Overall study status** | [_] Statistical analysis plan |
| 12/01/2024 | Completed | [X] Results |
| **Last Edited** | **Condition category** | [_] Individual participant data |
| 17/01/2025 | Nervous System Diseases | |

## Plain English summary of protocol

Background and study aims
Our brains generate electrical activity that can be measured through a process called electroencephalography (EEG), where electrodes are placed on the scalp. This method is commonly used in clinical settings, particularly for investigating brain diseases like epilepsy. Traditionally, EEG interpretation relies on visual analysis by experts, and there are concerns about the time pressures faced by reviewers and only moderate agreement among them. To address this, Holberg EEG has developed an EEG decision support tool using deep learning to aid in EEG interpretation, aiming to enhance agreement among reviewers. The aim of this study is to validate the SCORE-AI algorithm independently against an external gold standard derived from routine clinical assessments of electroencephalography (EEG) data at a center not involved in the algorithm's training. The SCORE-AI algorithm, developed by Holberg EEG, is an EEG decision support tool which uses deep learning to aid in EEG interpretation, aiming to enhance agreement among reviewers. The model distinguishes normal from abnormal recordings and then classifies the abnormal EEG into four main categories.

Who can participate?
Retrospective patient data obtained from the Montreal Neurological Institute and Hospital will be used for this study. Eligible EEG recordings will be selected by board-certified neurologists with a subspecialization in epilepsy and clinical neurophysiology from the Montreal Neurological Institute and Hospital.

What does the study involve?
This study involves analysing EEG data and the results from this study will not influence patient care.

What are the possible benefits and risks of participating?
There are no anticipated risks as this is a retrospective study and all data have been acquired in clinical routine. The results of the tests will be kept confidential, and EEG as well as clinical data will be shared completely de-identified with HolbergEEG for the specific purpose of this project. No identification of patients will be possible. The potential benefits of a positive outcome of this validation study include the possibility of implementing the SCORE-AI algorithm in diagnostics. This is particularly valuable in resource-limited settings lacking EEG experts, making diagnostic

capabilities more widely accessible. Additionally, in regions with expert availability, the algorithm could aid physicians in reducing their workload by providing efficient diagnostic support.

Where is the study run from?
Patient recruitment and clinical analysis will be conducted at the Montreal Neurological Institute and Hospital. AI analysis will be performed fully blinded by Holberg EEG. Data synthesis and analysis will be performed by an independent investigator not involved in clinical or AI analysis at the Montreal Neurological Institute and Hospital.

When is the study starting and how long is it expected to run for?
October 2022 to July 2024

Who is funding the study?
Investigator initiated and funded

Who is the main contact?
1. Dr Birgit Frauscher, birgit.frauscher@mcgill.ca
2. Dr Sandor Beniczky, sbz@filadelfia.dk

# Contact information

### Type(s)
Public, Scientific, Principal investigator

### Contact name
Prof Birgit Frauscher

### ORCID ID
https://orcid.org/0000-0001-6064-1529

### Contact details
3801 Rue University
Montreal
Canada
H3A 2B4
+1 (0)9196139386
birgit.frauscher@mcgill.ca

# Additional identifiers

### Clinical Trials Information System (CTIS)
Nil known

### Protocol serial number
2023-9123

# Study information

### Scientific Title

Evaluation of SCORE-AI for EEG classification versus the human expert

## Acronym
SCORE-AI

## Study objectives
1. The researchers aim to assess the accuracy of the SCORE-AI algorithm by comparing its performance to an external gold standard established through clinical evaluation of electroencephalograms (EEGs) from a separate center. Notably, this center, uninvolved in the algorithm's training, utilizes a different EEG equipment (Nihon Kohden) not employed during the algorithm's development. It is hypothesised that SCORE-AI will demonstrate excellent performance in categorizing EEGs into five main categories (normal, focal epileptiform, generalized epileptiform, focal non-epileptiform, and diffuse non-epileptiform) that is non-inferior to the consensus reached by expert scorers.
2. The researchers aim to evaluate the performance of SCORE-AI on extended EEG recordings from the epilepsy monitoring unit (20-hour) and compare it with its performance on shorter 25-minute EEG recordings. It is hypothesized that SCORE-AI will show identical performance metrics when applied to 20-hour EEG recordings compared to its performance on 25-minute EEG recordings.
3. The researchers aim to validate a novel feature in SCORE-AI which assesses the posterior dominant rhythm (PDR) in routine 25-minute recordings against the current gold standard consisting of visual analysis by a board-certified neurophysiologist. It is hypothesized that SCORE-AI will be non-inferior to the current gold standard.

## Ethics approval required
Ethics approval required

## Ethics approval(s)
approved 01/10/2022, Research Ethics Office of the Faculty of Medicine and Health (McGill University) (McIntyre Medical Building, #633-3655 Promenade Sir William Osler, Montreal, H3G 1Y6, Canada; +1 (0)5143983124; ilde.lepore@mcgill.ca), ref: 2023-9123

## Study design
Observational case-control study

## Primary study design
Observational

## Study type(s)
Other

## Health condition(s) or problem(s) studied
Epilepsy

## Interventions
Patient selection:
The researchers will retrospectively identify 104 EEG tracings from the MNI EEG hospital database, from consecutive patients between 20/10/2022 and 20/10/2016. This time frame coincides with the replacement of the EEG system with Nihon Kohden equipment. The selection process will continue until the desired patient targets are met: 50% of patients with a normal EEG, 12.5% with focal epileptiform anomalies, 12.5% with generalized epileptiform anomalies,

12.5% with focal non-epileptiform anomalies, and 12.5% with diffuse non-epileptiform anomalies. The researchers will first start to identify patients with normal EEG as potential bottleneck, and then in a second step select the proportions of the pathological categories as per the numbers above. The normal dataset consists of patients with a condition other than epilepsy such as e.g. psychogenic non-epileptic seizures, sleep disorders, or syncope as concluded based on the recording of the patients´ habitual clinical episodes. Based on previous results (Tveit et al., 2023), a sensitivity of 90% and a specificity of 88% is expected. If the study includes a total of 100 patients, there will be a 10% error in the 95% confidence interval. For 200 patients, it would be a 5% error.

Data segment selection:
For each patient, representative EDF files of 25-minute duration (either one routine EEG recording containing the abnormal grapho-element or a routine EEG and a 25-minute representative EEG file extracted from the epilepsy monitoring stay), which contain the main interictal abnormality corresponding to the gold standard (for example focal interictal spikes, when the seizure-diagnosis is focal seizure) or alternatively a routine EEG recording (for non-epileptiform abnormalities) as well as a 20-hour continuous EEG recording if available containing also the main interictal abnormality will be selected by one board-certified neurologist from the Montreal study team with subspecialization in epilepsy and clinical neurophysiology.

Analysis strategy:
The diagnostic gold standard in this study will be determined by the Montreal clinical team's summary report from the epilepsy monitoring unit stay. This report is based on the evaluation of patients' habitual paroxysmal events, considering both ictal EEG and semiology. For the categories of normal, focal epileptiform, and generalized epileptiform, the gold standard is derived from the patients' habitual episodes. For focal non-epileptiform and diffuse non-epileptiform categories, information is extracted from the medical chart, as these categories do not go along with paroxysmal events. This leads us to term it an "EXTERNAL gold standard" since it relies on seizures or clinical chart information, while the automatic model will primarily analyze interictal signals. It is widely accepted that analysis of the patients' habitual episodes is the most reliable diagnostic method for patients with paroxysmal episodes (epileptic or not). To balance the dataset, approximately 50% of the cases will have a condition other than epilepsy such as psychogenic non-epileptic seizures, sleep disorders, or syncope as concluded based on the recording of the patients' habitual clinical episodes. These patients will have normal interictal EEG. The other half of the cases will contain an approximately equal distribution of the four abnormal categories: focal epileptiform, generalized epileptiform, focal non-epileptiform, and diffuse non-epileptiform. The analysis will be performed fully blinded on the selected interictal EEG data segments by the external collaborators using SCORE-AI as a validated automatic model as well as by a consensus of three board-certified neurologists with subspecialization in epilepsy and clinical neurophysiology from the Montreal team that were not involved in data selection and are hence also fully blinded. Their consensus decision will serve for benchmarking (i.e. to answer the question: how is the automatic model performing compared with the experts?). Since the gold standard is external, we can even prove that the AI model is superior (note: when the majority consensus of the experts was used as the gold standard, it was impossible to assess the superiority of the AI model, because, by definition, one cannot have something better than the gold standard).

For all objectives, the researchers will calculate the accuracy, sensitivity, specificity, positive predictive value and negative predictive value of the algorithm using the clinical gold standard, and the same performance measures for the experts' scorings, using the clinical gold standard.

## Intervention Type

Other

### Primary outcome(s)
The accuracy, sensitivity, specificity, positive predictive value, and negative predictive value of the SCORE-AI algorithm in classifying the five main EEG categories will be compared to that of human experts. These variables will be assessed after all retrospective data have been extracted.

### Key secondary outcome(s))
1. The accuracy, sensitivity, specificity, positive predictive value, and negative predictive value of the SCORE-AI algorithm when used on long-term (20-hour) recordings will be assessed after all retrospective data have been extracted.
2. The accuracy, sensitivity, specificity, positive predictive value, and negative predictive value of the new SCORE-AI feature (posterior dominant rhythm) will be assessed in 25-minute routine EEG recordings. These variables will be assessed after all retrospective data have been extracted.

### Completion date
01/07/2024

## Eligibility

### Key inclusion criteria
1. Patients ≥15 years of age whose EEG classify as either normal, focal epileptiform, generalized epileptiform, focal non-epileptiform, or diffuse non-epileptiform

### Participant type(s)
Patient

### Healthy volunteers allowed
No

### Age group
Mixed

### Lower age limit
15 years

### Sex
All

### Total final enrolment
104

### Key exclusion criteria
1. Patients <15 years of age
2. Dual pathology
3. Absence of sufficient clinical information to grade in 1 of the 5 categories such as the absence of recorded seizures during the stay in the epilepsy monitoring unit or insufficient information on etiology of non-epileptiform EEG anomalies from the medical chart

### Date of first enrolment

20/10/2016

**Date of final enrolment**
20/10/2022

# Locations

**Countries of recruitment**
Canada

Denmark

Norway

**Study participating centre**
**Montreal Neurological Institute and Hospital**
3801 University Street
Montreal
Canada
H3A 2B4

**Study participating centre**
**Danish Epilepsy Centre Filadelfia**
Kolonivej 1
Dianalund
Denmark
4293

**Study participating centre**
**Holberg EEG AS**
Fjøsangerveien 70 A
Bergen
Norway
5068

# Sponsor information

**Organisation**
McGill University

**ROR**

https://ror.org/01pxwe438

# Funder(s)

**Funder type**
Other

**Funder Name**
Investigator initiated and funded

# Results and Publications

**Individual participant data (IPD) sharing plan**
The features necessary to reproduce the study's results will be shared through a link provided in the final publication, accessible via the lab's Github repository. However, note that the raw individual participant data obtained for this project is not expected to be made available in the repository. This decision aligns with the Institutional Review Board approval obtained for the project, which did not encompass open data sharing.

**IPD sharing plan summary**
Not expected to be made available

**Study outputs**

| Output type | Details | Date created | Date added | Peer reviewed? | Patient-facing? |
|---|---|---|---|---|---|
| Results article | | 14/08/2024 | 17/01/2025 | Yes | No |