Evaluation of autoSCORE: an artificial intelligence based algorithm for EEG classification versus human experts

Document Name	Study Protocol
Public Title	Evaluation of autoSCORE: an artificial intelligence-based algorithm for EEG classification versus human experts
Scientific Title	Accuracy of EEG classification by autoSCORE algorithm compared with human experts
Acronym	autoSCORE
Document Number	24084-02
Version	4.0
Release Date	28 Feb 2022

Abbreviations

Abbreviation	Explanation
EEG	Electroencephalography
HE	Human expert
HUS	Haukeland University Hospital, Norway
OUS	Oslo University Hospital, Norway
FEH	Filadelfia Epilepsy Hospital, Denmark
SCORE	Standardized Computer-based Organized Reporting of EEG
VM	Virtual machine

STUDY DESCRIPTION

Source of Monetary and Material Support
The study is funded by Holberg-EEG AS (Fjøsangerveien 70 A, 5068 Bergen, Norway)
Web: holbergeeg.com
Phone: +47 926 44 261

Principal Investigator
Professor Sándor Beniczky (Danish Epilepsy Centre and Aarhus University Hospital, Denmark) takes responsibility for initiating and managing the study.

Contact for Public and Scientific Queries
Professor Sándor Beniczky, MD, PhD,
Danish Epilepsy Centre and Aarhus University
Address: Epilepsihospitalet Filadelfia, Visby Allé 5, 4293 Dianalund, Denmark
Phone: +4526981536
Email: sbz@filadelfia.dk

Countries of Recruitment
Denmark, Norway, USA.

Problems Studied
Electroencephalography (EEG) in patients suspected for epilepsy, seizures, impaired consciousness or altered cognition.

INTERVENTIONS

Background

Electroencephalography (EEG) measures electric brain activity using electrodes attached to the scalp. This is used in clinical practice to investigate brain disease, most commonly epilepsy, coma, and dementia. The clinical interpretation of EEGs is until now mainly based on expert visual analysis (Tatum IV et al. 2016), and there are indications that EEG reviewers are under increasing time pressures (Ng and Gillis 2017; Brogger et al. 2018). The interrater agreement assessing EEG studies is only moderate (Van Donselaar et al., 1992; Stroink et al., 2006). Holberg EEG has initiated the development of an EEG decision support tool based on deep learning techniques with the purpose of assisting the process of EEG interpretation and increase the interrater agreement. Hospital partners at Haukeland University Hospital (Norway), Filadelfia Epilepsy Hospital (Denmark), and Oslo University Hospital have for many years used SCORE-EEG software developed by Holberg EEG to assess and tag EEG in a standardized way,

and at the same time produce a large database of tagged EEGs. This database is used to train an algorithm (autoSCORE) to automatically assess EEGs. autoSCORE will be trained to separate normal from abnormal EEGs. When autoSCORE assesses the EEG as abnormal it will further sub-classify abnormalities into one or more of the subgroups focal epileptiform abnormality, generalized epileptiform abnormality, focal non-epileptiform abnormality, and diffuse non-epileptiform abnormality.

Objective

To evaluate the accuracy of autoSCORE in distinguishing between normal and abnormal EEG recordings, and classifying the abnormal EEG recordings into the four major clinical categories: focal-epileptiform, generalized-epileptiform, diffuse-slowing (non-epileptiform), focal-slowing (non-epileptiform).
In this is a diagnostic accuracy study, index-test is autoSCORE, and reference standard is evaluation of the routine EEG recordings by HEs. In the phase-3 part of the study, reference standard is the majority consensus of a panel of 11 HEs. In the phase-4 part of the study, reference standard is the clinical assessment of the EEGs, as part of the routine, by HEs at a centre which did not participate in the development of autoSCORE.

Methods

- **Inclusion and exclusion criteria**:
  - *Inclusion*: Routine clinical EEG recordings in patients referred to EEG on suspicion of epilepsy or seizures, and patients referred to EEG on for diagnostic work-up in patients with impaired consciousness or cognitive impairment.
  - *Exclusion*: Patients younger than 3 months, and critically ill patients with rhythmic or periodic EEG patterns.
- **Index test**: AutoSCORE analysis of the EEG recordings. The analysis is fully automated and blinded to all other data. The algorithm and the detection threshold values are fixed (pre-defined according to the previous development process). No iterations are allowed.
- **EEGs in phase-3**: The EEGs to be included into this study have not been part of the training dataset to develop the autoSCORE algorithm. The routine clinical EEGs are recorded at HUS, FEH and at Mayo Clinics. The distribution in this representative validation dataset should be as follows:

	Pediatric **<16 years**	**Adult** **>16 years**	**Row sum**
Normal EEG	15	28	43
Abnormal EEG	20	37	57
Column sum	35	65	100

With the above described distribution, 75 EEGs will be randomly selected from the independent test-datasets of 3.000 EEGs from HUS and FEH and 25 EEGs from the

independent test-dataset of 140 EEGs from the Mayo Clinics. All EEGs will be anonymized by the Hospitals before they are transmitted to Holberg. Security of assessments will be assured by restricting access of the HE to their own Excel sheet for storing their assessments, which they can edit with the data in the predetermined columns. Once complete, the HE will sign the Excel sheet and send to Holberg EEG for placement on the SharePoint site. Holberg EEG is blinded to the HE assessments, until the autoSCORE results are documented for all EEGs.

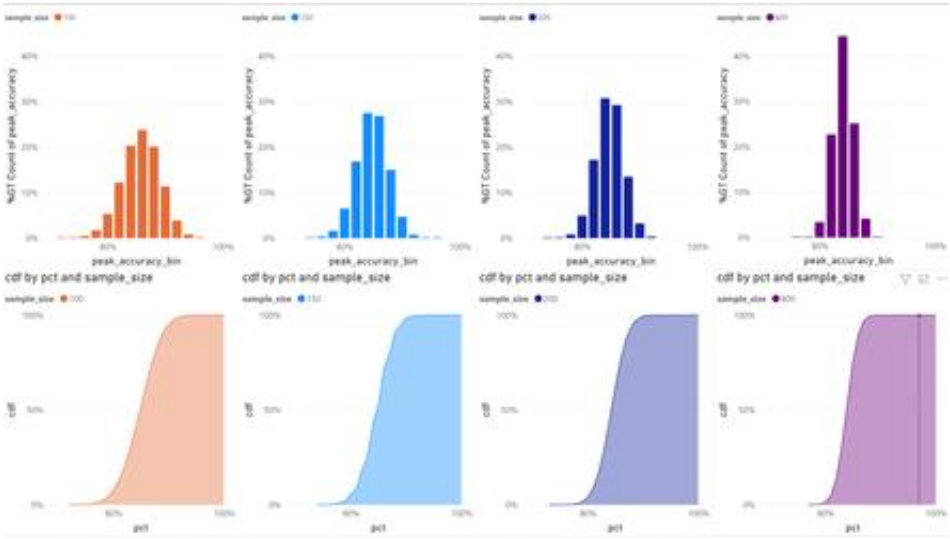
- **Reference standard in phase-3**: Majority consensus of a panel of HEs, who assess independently 100 routine clinical EEGs. Each EEG is assessed by 11 HEs, who will make the following decisions:
    - EEG is normal or abnormal
    - If the EEG is abnormal, HEs assess if one or more of the following categories of abnormality is present:
        - focal-epileptiform abnormality
        - generalized epileptiform abnormality
        - focal-slowing (non-epileptiform) abnormality
        - diffuse-slowing (non-epileptiform) abnormality

  HEs are blinded to autoSCORE.
- **EEGs in phase-4**: 9,785 consecutive EEG recordings from OUH, fulfilling inclusion and exclusion criteria. These recordings have not been used to train the algorithm, and this centre did not participate in development of the algorithm.
- **Reference standard in phase-4**: clinical EEG assessment of the recordings, by HEs evaluating these EEGs as part of the patients´ routing diagnostic workup. The HE assessment is blinded to autoSCORE. Fourteen HEs contributed to the clinical EEG assessment of the EEGs included into phase-4.
- **Benchmarking**: Currently there isn´t any commercially available or published algorithm which provides a comprehensive, fully automated assessment of routine, clinical EEG recordings, comparable with autoSCORE. However, the ENCEVIS software (FDA approved) has a functionality for automated detection of epileptiform discharges. This corresponds to a combination of two of the four categories in the classification of EEG abnormalities (focal-epileptiform and generalized-epileptiform). We will compare the accuracy of autoSCORE and ENCEVIS to identify these combined classes.
- **Outcome measures**:
    - *Primary outcome measures*: diagnostic accuracy parameters, according to the STARD criteria. We will calculate: sensitivity, specificity, accuracy, positive predictive value, negative predictive value and F1-score, for the EEGs in the phase-3 part of the study.
    - *Secondary outcome measures*: Inter-test agreement (autoSCORE vs. HE) in the phase-4 part of the study.

Sample Size

Simulations showed the random distribution of measured accuracy for sample sizes of 100, 150, 200 and 400 recordings in the training dataset. The simulations showed the accuracy was not significantly increased with the higher sample sizes. The simulation is based on binary classification. For sub-classification, similar results can be expected of a similar level of

accuracy reached (if it is less accurate, then the random variation increases). The diagram illustrates this. To make the process feasible for the HEs, considering that the process of visual evaluation is time-consuming, we need to limit the size of the phase-3 validation dataset to the lowest representative number. Therefore, HEs, need to assess 100 EEGs.



Tools and procedures

Excel has been selected for use by the participants as it is easily accessible and generally well understood. SharePoint has been selected as it is an easily managed tool that meets the needs of accessibility while maintaining the maintaining the integrity of the study. A Virtual Machine is set up to host the NeuroWorks EEG software (version 9.2.0.6628-54426). The number of human experts need to be at least seven. Previous studies on inter-rater variability in EEG showed that majority consensus of a panel of human experts does not change significantly beyond seven raters. There will be an even distribution of HEs from North America and Europe. All the HEs are board certified in Clinical Neurophysiology, or hold specialty competence within Clinical Neurophysiology or Neurology including EEG reading competence.

Instructions for Human Experts

The HE will get instructions for how to:

- Open the virtual machine (VM) where the necessary infrastructure is set up for each individual HE.
- How to operate the EEG software
- Subgroup definitions
- How to report the assessments of each EEG in an Excel sheet installed at the VM.
- How to send a screenshot of the finalized Excel sheet to Holberg when all EEGs are assessed.

EEG Data Provision

All EEGs have been provided to the study under a legal contract with the relevant institution, which have been responsible for anonymization of the data, which has removed the need for individual patient consent.

Data Evaluation

1. The Excel sheet has been set up with data validation to ensure that only relevant data are inserted.
2. The Excel sheet has been used to prevent editing by HE of cells that has already been prefilled by Holberg.
3. After HE has finalized all their assessments in the Excel sheet, they are instructed to take a screenshot and send this to Holberg.
4. The HEs will also be send a wet signed copy of the final assessment sheet.
5. The SharePoint and dedicated inbox will be monitored by the Clinical & RA Manager.

Overall trial start date

June 1st, 2021.

Ethics Review

IRB and data safety approval.

Reference number: “Sagsnr. 0100256”. Date: July 7th 2020

Contact details: Pernille Worm (legal counsel, DPO) Direktionssekretariatet, Kolonivej 1, st., 4293 Dianalund. Phone: 58264200. Email: pwo@filadelfia.dk

IPD sharing statement

Individual clinical trial participant-level data (IPD) will be shared upon request.

Contact: Professor Sandor Beniczky, Danish Epilepsy Centre and Aarhus University, Denmark (Visby Allé 5, 4293 Dianalund, Denmark; Phone:+4526981536; Email: sbz@filadelfia.dk).

Type of data: For the phase-3 dataset the anonymised EEG, Diagnostic Gold standard; Demographics (age, gender), output of the algorithm will be available upon request.

Data will be available upon request, for 10 years from the publication, for scientific non-commercial use. As the dataset is de-identified, there is no need for consent from the participants.

References

1. Beniczky, Sándor, Harald Aurlien, Jan C Brøgger, Anders Fuglsang-Frederiksen, António Martins-da-Silva, Eugen Trinka, Gerhard Visser, Guido Rubboli, Helle Hjalgrim, and Hermann Stefan. 2013. "Standardized computer-based organized reporting of EEG: SCORE." Epilepsia 54 (6): 1112-1124.
2. Beniczky, Sándor, Harald Aurlien, Jan C Brøgger, Lawrence J Hirsch, Donald L Schomer, Eugen Trinka, Ronit M Pressler, Richard Wennberg, Gerhard H Visser, and Monika Eisermann. 2017. "Standardized computer-based organized reporting of EEG: SCORE–second version." Clinical Neurophysiology 128 (11): 2334-2346.
3. Brogger, Jan, Tom Eichele, Eivind Aanestad, Henning Olberg, Ina Hjelland, and Harald Aurlien. 2018. "Visual EEG reviewing times with SCORE EEG." Clinical neurophysiology practice 3: 59-64.
4. Ng, Marcus C, and Kara Gillis. 2017. "The state of everyday quantitative EEG use in Canada: a national technologist survey." Seizure 49: 5-7.
5. Stroink, Hans, Robbert-Jan Schimsheimer, Al W de Weerd, Ada T Geerts, Willem F Arts, Els A Peeters, Oebele F Brouwer, A Boudewijn Peters, and Cees A van Donselaar. 2006. "Interobserver reliability of visual interpretation of electroencephalograms in children with newly diagnosed seizures." Developmental medicine and child neurology 48 (5): 374-377.
6. Tatum IV, William O, Olga Selioutski, Juan G Ochoa, Heidi Munger Clary, Janna Cheek, Frank W Drislane, and Tammy N Tsuchida. 2016. "American clinical neurophysiology society guideline 7: guidelines for EEG reporting." The Neurodiagnostic Journal 56 (4): 285-293.
7. van Donselaar, Cees A, Robert-Jan Schimsheimer, Ada T Geerts, and August C Declerck. 1992. "Value of the electroencephalogram in adult patients with untreated idiopathic first seizures." Archives of neurology 49 (3): 231-237.
8. HEEG-autoSCORE-IU-01, Intended Use Record autoSCORE Revision 3.0