1.6 Overall synopsis of the clinical investigation

Title	Performance of BIOSERENITY AI encephalopathy algorithm in classifying encephalopathic EEG automatically: a retrospective validation study.
Short title	BIOSERENITY AI ENCEPHALOPATHY EEG validation study
Lay title, if applicable	Not applicable
CIP version and date	Version 1.0 – 13/10/2026
EUDAMED Single Reference Number (SRN) or CIV-ID, if previously assigned	FR-MF-000039391
CI modification number, if applicable	Version 1.0
Sponsor name and address	BIOSERENITY MEDICAL DEVICES GROUP 6/8 rue Jean Antoine de Baïf, 75013 Paris
Participating Location(s) and country(ies)	Data were collected from several hospitals in France through Neurophy, the BioSerenity tele-interpretation service.
Name of Investigational Device	BIOSERENITY AI EEG ENCEPHALOPATHY
Clinical investigation Purpose and Background • Rationale for CI • Background of device and condition • Current standard of care	An electroencephalogram (EEG) exam, in conjunction with clinical evaluation, is frequently used in neurological investigations but requires very specific expertise. Because of this need for specific expertise, Artificial Intelligence (AI) can help with diagnosis and democratize the use of EEGs. Several papers have been published to address this classification challenge using Artificial Intelligence. The BIOSERENITY AI EEG ENCEPHALOPATHY is an advanced algorithm designed to automate the classification of encephalopathies in EEG records. Classifying encephalopathic records could broaden the use of EEG for a better and quicker diagnosis in critical cases. The clinical validation study aims to confirm the performance of the BIOSERENITY AI EEG ENCEPHALOPATHY algorithm, ensuring it meets or exceeds the standards set by currently marketed devices.
Study device and Comparator	**Study device and Comparator** *Study device* BIOSERENITY AI EEG ENCEPHALOPATHY is the second stage of a two-stage sequential pipeline of algorithms.

Stage	Algorithm	Input	Predictions (output scores ; s=thresholds)
1	Normal/Abnormal (CE-marked, 12/2025)	Full standard EEG record (~20 min)	Normal ($s \leq 0.22$), Uncertain ($0.22 < s \leq 0.46$), Abnormal ($s > 0.46$)
2	Encephalopathy (investigational algorithm)	Records predicted Abnormal by Stage 1	Non-encephalopathy ($s < 0.35$), Uncertain ($0.35 \leq s \leq 0.64$), Encephalopathy ($s > 0.64$)

Both stages use fixed decision thresholds. These thresholds were calibrated on an independent test dataset with no overlap with the validation database, and are frozen for this study.

Reference standard: physician annotations

Three qualified physicians independently annotate the same EEG records:

- one physician annotates in routine clinical practice;
- two physicians annotate retrospectively, blinded to the routine annotation, to each other and to all algorithm predictions.

Each physician assigns one of three labels:

- **IA1: certainly normal.** No pathological activity and a normal background with alpha rhythm. Drug impregnation, sedation or a fast rhythm may be present if the trace is still read as normal.
- **IA2: undetermined** (exceptional). Excessive artifacts, or impregnation or sedation preventing any conclusion.
- **IA3: certainly abnormal.** Physicians specify a sub-label:
    - IA3L: lesional;
    - IA3E: diffuse encephalopathy;
    - IA3S: status epilepticus;
    - IA3Ep: epileptiform activity;
    - IA3A: other abnormality

The **consensus annotation** is the sub-label assigned by at least 2 of the 3 physicians. The following records are excluded from the validation database:

- records without a 2/3 consensus at the sub-label level;
- records with a consensus annotation of IA2;
- records whose consensus annotation carries more than one sub-label (e.g. IA3E with IA3S), since encephalopathy may coexist with other abnormalities such as status epilepticus.

The number of records excluded for each reason is reported.

	Performance analyses **Cascade (Stage 1 + Stage 2).** o Input: all included records. o Positive class: consensus IA3E (diffuse encephalopathy). o Negative class: consensus IA1 or any IA3 sub-label other than IA3E. o Records predicted Normal by Stage 1 count as negative predictions. Stage 2 runs only on records predicted Abnormal by Stage 1. Records predicted Uncertain, at either stage, are excluded from the performance computation. The proportion of Uncertain predictions is reported for each stage and each analysis. *Comparator* Performance of BIOSERENITY AI EEG ENCEPHALOPATHY is compared with that reported for a similar marketed device, DeltaScan (510(k) K222680).
Clinical development stage	Pivotal study
Design of the clinical investigation	Retrospective, non-interventional, multicenter, validation study
Objectives	Primary objective • To evaluate the classification performance of BIOSERENITY AI EEG ENCEPHALOPATHY against a consensus of physicians versus the performance of a similar device (DeltaScan). Secondary objectives • To evaluate the classification performance of the algorithm on sensitivity, specificity and F1 score. • To evaluate the classification performance of the algorithm (cascade) according to patients' demographic data. • To evaluate the distance algorithm-consensus versus distance physicians - consensus. • To evaluate the classification performance of the algorithm compared to the scientific literature. • To evaluate the classification performance of the algorithm according to pathology category (lesional, diffuse encephalopathy, status epilepticus, epileptiform activity, undetermined).
Primary endpoints	Evaluation of negative predictive value and sum of "negative predictive value + positive predictive value" of the classification of all the records on the consensus of physicians versus the performance of a similar device (DeltaScan). • Negative predictive value is a measure of how often a negative test result is actually negative • Positive predictive value is a measure of how often a positive test result is actually positive

Secondary endpoints	• Evaluation of classification performance of the algorithm on sensitivity, specificity and F1 score.
	  • Sensitivity (true positive rate) is the probability of a positive test result, conditioned on the individual truly being positive.
	  • Specificity (true negative rate) is the probability of a negative test result, conditioned on the individual truly being negative.
	  • F1 score is the harmonic mean of sensitivity and positive predictive value for the encephalopathy class.
	• Evaluation of classification performance of the algorithm depending on patients' demographic data.
	• Evaluation of the distance algorithm-consensus versus distance physicians - consensus.
	• Evaluation of the classification performance of the algorithm on scientific literature: computation of sensitivity, specificity, and F1 score.
	• Evaluation of the classification performance of the algorithm according to pathology category (lesional, status epilepticus, epileptiform activity, undetermined).

Description of participants / study population	The study population includes adult patients who performed an EEG for neurological clinical examination for several indications such as epilepsy, dementia, coma, stroke. 456 EEG records will be included, i.e. the validation database of the Stage 1 study, consistent with the 434 patients included in the 510(k) of the similar device DeltaScan.
Inclusion criteria	1. ≥18 years 2. Male or female 3. Having completed a standard EEG record (approx. 20min recording) 4. Individually informed and non-opposed for the use of their personal data for the conduct of the research 5. Neuronaute tracings only (not necessarily Icecap only) 6. Tracings for which the interpreter has indicated, after visual analysis of the EEG trace, one of the codes: "IA1", "IA2", "IA3L", "IA3E", "IA3S", "IA3Ep", "IA3A" 7. Records with double-blinded manual scoring by three (3) physicians.
Exclusion criteria	1. Tracings without a 2/3 consensus at the sub-label level are excluded. 2. Tracings with a consensus annotation of IA2 are excluded. 3. Tracings whose consensus annotation carries more than one sub-label (e.g. IA3E with IA3S) are excluded. 4. EEG lacking the 19 channels of the 10/20 positioning or for which some channels are non-functional / too artifactual.
Sample size	The objective of the study is to evaluate the performance of the study device in detecting and classifying encephalopathies in EEG records. As this study is the second stage of the validation of the two-stage pipeline, it reuses the validation database of the Stage 1 study (CE-marked), whose size is equivalent to the sample size reported for the similar device DeltaScan in its 510(k): 456 EEG

	records will be used, before application of the exclusion criteria specific to this study.
Duration and follow up of the clinical investigation	EEG records were collected from 22 May 2025 to 31 August 2025 (validation database of the Stage 1 study). The total expected duration of the study is approximately 2 to 3 months.
Statistical considerations	**Primary endpoint analysis** Computation of negative predictive value and sum of "positive predictive value + negative predictive value" is performed by considering the consensus of physicians as reference, and 95% confidence intervals (CI) are computed by bootstrap resampling (1000 resamples). Results will be compared to the performance of a similar device (DeltaScan): negative predictive value (>0.84 [0.79-0.88]) and sum of negative predictive value and positive predictive value (>1.63 [1.55-1.71]). Results will be compared to the performance claimed by the similar device DeltaScan in its 510(k) summary (K222680) with a 10% margin. **Secondary endpoint analyses** • Specificity, sensitivity, F1 score with their 95% CI computed by bootstrap resampling (1000 resamples). • Negative predictive value, positive predictive value, sensitivity, specificity and F1 score with their 95% CI computed by bootstrap resampling (1000 resamples) for each demographic group (age and sex). Sensitivity (IA3E) and specificity (other categories) with their 95% CI for each pathology category (normal, lesional, status epilepticus, epileptiform activity, other abnormality). • To evaluate the classification performance of the algorithm on scientific literature by computing sensitivity, specificity, and F1 score whose 95% confidence interval is computed by bootstrap resampling (1000 resamples).