
C O N F I D E N T I A L

Performance de l'algorithme BIOSERENITY A.I. dans la classification de l'EEG comme normal ou anormal

Plan Investigation Clinique:	Version 1.0
Date:	15 MAI 2025
Titre court:	BIOSERENITY A.I EEG NORMAL/ANORMAL – étude de validation
Indication	Classification des EEG normaux et anormaux à l'hôpital
Méthodologie:	Rétrospective, non-interventionnelle, multicentrique
Pays	France
Promoteur:	BIOSERENITY MEDICAL DEVICES GROUP 6/8 rue Jean Antoine Baïf, 75013 Paris
Coordinateur scientifique:	Professeur DERAMBURE Philippe

Ce document est une communication confidentielle de Bioserenity. Les informations qu'il contient ne peuvent être reproduites ou diffusées d'une autre manière sans l'accord de Bioserenity.
--

Résumé

Titre	Performance de l'algorithme BIOSERENITY A.I. dans la classification automatique des EEG normaux-anormaux : une étude de validation rétrospective.
Titre court	BIOSERENITY A.I. EEG NORMAL/ANORMAL - étude de validation.
Version et date	Version 1.0– 15/05/2025
EUDAMED Single Reference	Pas applicable
Coordonnées du Promoteur	BIOSERENITY MEDICAL DEVICES GROUP 6/8 rue Jean Antoine de Baïf, 75013 Paris
Lieu(x) et pays participant(s)	Les données seront recueillies auprès d'une soixantaine d'hôpitaux en France par l'intermédiaire de Neurophy, le service de télé-interprétation de BioSerenity.
Nom du dispositif médical	BIOSERENITY A.I. EEG NEURO
Objectif et Contexte	L'électroencéphalogramme (EEG), en conjonction avec l'évaluation clinique, est fréquemment utilisé dans les examens neurologiques, cependant il nécessite une expertise médicale très spécifique. En raison de ce besoin d'expertise spécifique, l'Intelligence Artificielle (I.A.) pourrait aider les professionnels de santé dans leurs diagnostics et démocratiser l'utilisation de l'EEG dans les hôpitaux. Plusieurs articles ont été publiés dans ce sens pour répondre à ce défi de classification automatique des enregistrements à l'aide de l'I.A. BIOSERENITY A.I. EEG normal/anormal est un algorithme avancé conçu pour automatiser la classification des enregistrements EEG comme normaux ou anormaux. Cette classification automatique pourrait élargir l'utilisation de l'EEG pour un diagnostic facilité et plus rapide dans les cas critiques. L'étude de validation clinique a pour but de confirmer la performance et la sécurité de l'algorithme A.I. BIOSERENITY EEG normal/anormal, en s'assurant qu'il respecte ou dépasse les normes fixées pour les dispositifs médicaux.
Dispositif à l'étude et comparateur	La validation clinique de l'algorithme sera basée sur sa capacité à classer les enregistrements EEG en deux catégories : • “IA 1” pour un tracé “certainement normal”: ○ Absence d'activité pathologique ○ Activité de fond normale avec rythme alpha ○ Imprégnation médicamenteuse (ex : benzodiazépines), mais la lecture permet de conclure que le tracé est normal. La présence d'un rythme rapide n'empêche pas de conclure.

	○ Tracé de sédation, mais la lecture permet de conclure que le tracé est normal.. • “IA 3” pour un tracé “certainement anormal” précisant les différentes catégories de pathologies : ○ IA3L: Lésionnel (activité pathologique focale, indication d’une imagerie). ○ IA3E: Encéphalopathie diffuse. ○ IA3S: Etat de mal épileptique (focal ou généralisé). ○ IA3A: Autres/indéterminés. ○ IA3Ep: Présence d'une activité épileptique. Cette classification de l’algorithme sera comparée à la notation fournie par des médecins qualifiés qui ont classé les enregistrements en aveugle (3 médecins par enregistrement afin d'obtenir un consensus : 1 en pratique clinique courante / 2 en aveugle). Elle sera utilisée comme référence pour évaluer la performance des algorithmes. Le consensus sera obtenu par une notation à la majorité des 2/3, définie comme une notation manuelle par un consensus entre trois (3) médecins en double aveugle. Les enregistrements EEG pour lesquels il n'y a pas de consensus sont exclus de la base de données de validation finale. Pour les médecins, une troisième catégorie “IA 2” qui regroupe les enregistrements EEG pour lesquels il est impossible de trancher entre certainement normal et certainement anormal (cotation exceptionnelle) est ajoutée si : • Présence de trop d’artefacts • Imprégnation médicamenteuse ou tracé sous sédation empêchant toute conclusion (présence excessive de rythme rapide)
Phase de développement clinique	Etude pivot
Design de l’étude	Rétrospective, non interventionnelle, multicentrique
Objectifs	<u>Objectif primaire</u> • Évaluer la performance de classification de l'algorithme BIOSERENITY A.I. EEG normal/anormal par rapport à la classification d'un consensus de neurologues versus la performance d’un dispositif similaire (autoSCORE). <u>Objectifs secondaires</u> • Evaluer la performance de classification de l’algorithme selon les données démographiques des patients. • Evaluer la variabilité inter-scorer • Evaluer la distance algorithme-consensus versus distance neurologues – consensus • Evaluer les faux négatifs et les faux positifs renvoyés par l’algorithme. • Evaluer la performance de classification de l’algorithme sur les données publiques TUAB. • Evaluer la performance de classification de l’algorithme en comparaison à la littérature scientifique.

Critère de jugement primaire	Évaluation de la sensibilité, de la spécificité, de la précision équilibrée, des valeurs prédictives positives et négatives de la classification de tous les enregistrements sur le consensus de neurologues par rapport à la performance d'un dispositif similaire (autoSCORE). • La sensibilité (taux de vrais positifs) est la probabilité d'un résultat positif, à condition qu'il soit réellement positif. • La spécificité (taux de vrais négatifs) est la probabilité d'un résultat négatif, à condition qu'il soit réellement négatif. • La précision équilibrée est la moyenne du rappel obtenu dans chaque classe. • La valeur prédictive positive est une mesure de la fréquence à laquelle un résultat de test positif est réellement positif. • La valeur prédictive négative est une mesure de la fréquence à laquelle un résultat négatif est réellement négatif.
Critères de jugement secondaires	• Evaluer la performance de classification de l'algorithme selon les données démographiques des patients • Evaluation des résultats du consensus de neurologues par: ○ Détermination de la variabilité inter-scorer ; ○ Détermination de la distance algorithme-consensus versus distance neurologues – consensus. • Evaluation des faux négatifs et des faux positifs renvoyés par l'algorithme. • Evaluation de la performance de classification de l'algorithme sur les données TUAB et la littérature scientifique : calcul de la sensibilité, la spécificité, la précision (pour TUAB), et des valeurs prédictives positives/négatives comme mentionné pour le critère primaire.
Description de la population étudiée	La population de l'étude comprend des patients adultes dont l'EEG a été enregistré dans le cadre d'un examen clinique neurologique pour diverses indications telles que l'épilepsie, la démence, le coma et l'accident vasculaire cérébral. Environ 400 patients seront inclus dans l'étude dans l'objectif d'obtenir 350-400 enregistrements EEG exploitables.
Critères d'inclusion	1. ≥ 18 ans 2. Homme ou Femme 3. Ayant un enregistrement EEG standard (approx. 20min d'enregistrement) 4. Informés et non-opposés à l'utilisation de leurs données personnelles dans le cadre de l'étude 5. Tracés Neuronaute uniquement 6. Tracés pour lesquels l'interpréteur a indiqué après analyse visuelle de l'enregistrement EEG, un des codes "IA 1", "IA2" or "IA 3".

Critères d'exclusion	1. Tracés pour lesquels l'interpréteur n'a pas indiqué un des codes "IA 1", "IA2" or "IA 3". 2. EEG dépourvu des 19 canaux du système de positionnement 10/20 ou dont certains canaux ne sont pas fonctionnels / trop artéfactés.
Taille de l'échantillon	L'objectif de l'étude est d'évaluer la performance du dispositif d'étude dans la détection et la classification de l'EEG comme normal ou anormal. En raison de la grande hétérogénéité des pathologies et des populations qui seront étudiées et analysées, environ 350-400 enregistrements EEG seront nécessaires.
Durée de l'étude	La période de collecte des données s'étend du 16 Mai 2025 au 31 Août 2025.
Statistiques	<u>Analyse du critère d'évaluation primaire</u> Calcul de la spécificité (>84.6% [81.1-87.9]), sensibilité (>90.5% [87.6-93.2]), précision équilibrée (>87.6%), valeur prédictive positive (>85.2% [81.9-88.4]), valeur prédictive négative (>90.1% [87.0-92.9]) dont l'intervalle de confiance (IC) à 95% est calculé grâce à un ré-échantillonnage par bootstrap (1000 rééchantillons) sur le consensus de neurologues versus la performance d'un dispositif similaire (autoSCORE). <u>Analyse des critères d'évaluation secondaire</u> • Spécificité, sensibilité, précision (pour TUAB), valeur prédictive positive et valeur prédictive négative avec leur IC à 95 % calculés par rééchantillonnage bootstrap (1000 rééchantillons) pour chaque paramètre démographique. • Évaluer les résultats du consensus des neurologues en calculant le kappa de Cohen et son IC à 95 % avec un rééchantillonnage bootstrap (1000 rééchantillons). Les IC seront comparés : s'ils ne se chevauchent pas, la différence entre les accords est considérée comme statistiquement significative. • Détermination de la variabilité inter-scorer (>0.72) ([0.649, 0.796]). • Détermination de la distance algorithme-consensus (>0.9) ([0.820, 0.987] versus distance neurologue – consensus. • Evaluation des faux négatifs et des faux positifs renvoyés par l'algorithme : ils seront examinés par des neurologues afin de comprendre les raisons de l'échec de l'algorithme. • Evaluation de la performance de classification de l'algorithme sur les données TUAB et la littérature scientifique en calculant la sensibilité (>80.43%), spécificité (>93.27%), et la précision (>85.59%) (pour TUAB), dont l'intervalle de confiance à 95 % est calculé par rééchantillonnage bootstrap (1000 rééchantillons).