



Emerging Technologies in Epilepsy Care Across the Lifespan

**Chusak Limotai, MD., Ph.D.
Chula Epilepsy**



**" Clinicians
will not be replaced by ML/AI,
but clinicians without ML/AI will be replaced
by clinicians supported by ML/AI "**



" While explainable ML/AI is attractive, one benefit of ML/AI methods is to identify patterns across thousands of data points that humans cannot reasonably comprehend. "



Talk Overview

- **Diagnosis and Detection of Seizures and Other patterns**
- **Tools for Epilepsy Surgery**
- **Large Language Models (ChatGPT)**
- **Ethical and Practical Challenges**



Diagnosis and Detection of Seizures and Other Patterns

AI-assisted EEG interpretation



Automated Interpretation of Clinical Electroencephalograms Using Artificial Intelligence

Jesper Tveit, PhD; Harald Aurlen, MD, PhD; Sergey Plis, PhD; Vince D. Calhoun, PhD; William O. Tatum, DO;
Donald L. Schomer, MD; Vibeke Arntsen, MD; Fieke Cox, MD, PhD; Firas Fahoum, MD; William B. Gallentine, DO;
Elena Gardella, MD, PhD; Cecil D. Hahn, MD; Aatif M. Husain, MD; Sudha Kessler, MD;
Mustafa Aykut Kural, MD, PhD; Fábio A. Nascimento, MD; Hatice Tankisi, MD, PhD; Line B. Ulvin, MD;
Richard Wennberg, MD, PhD; Sándor Beniczky, MD, PhD

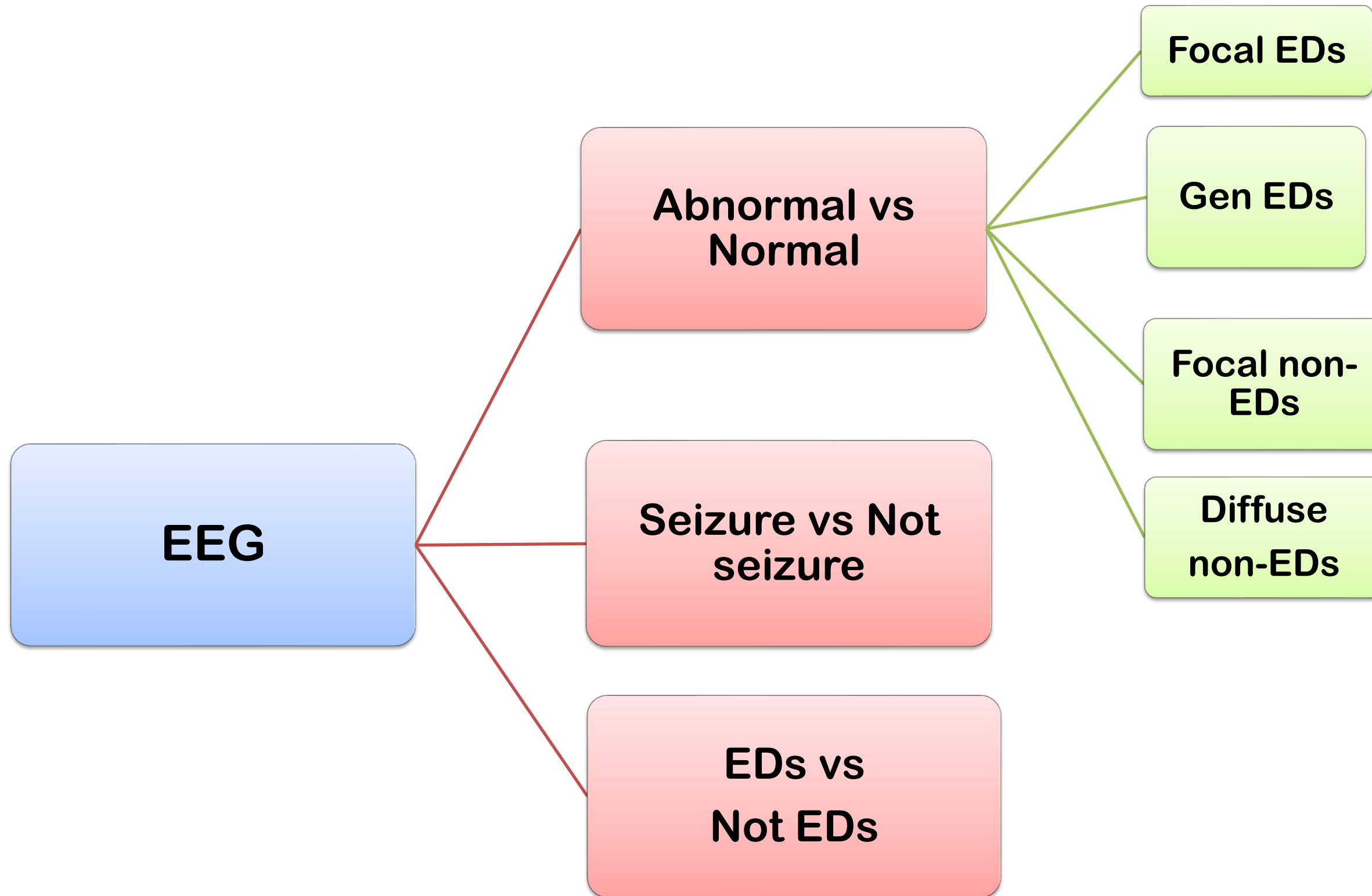
Standardized Computer-based Organized Reporting of EEG–Artificial Intelligence [**SCORE-AI**]

- Multicenter
- Convolutional neural network (CNN) model
- **30,493** recordings of patients referred for EEG were included into the development data set annotated by 17 experts
- Validated using 3 independent test data sets

Patients aged more than 3 months
and **Not critically ill were eligible**

(AutoScore®)

Ambulatory setting (Routine EEG at OPD)



SCORE-AI

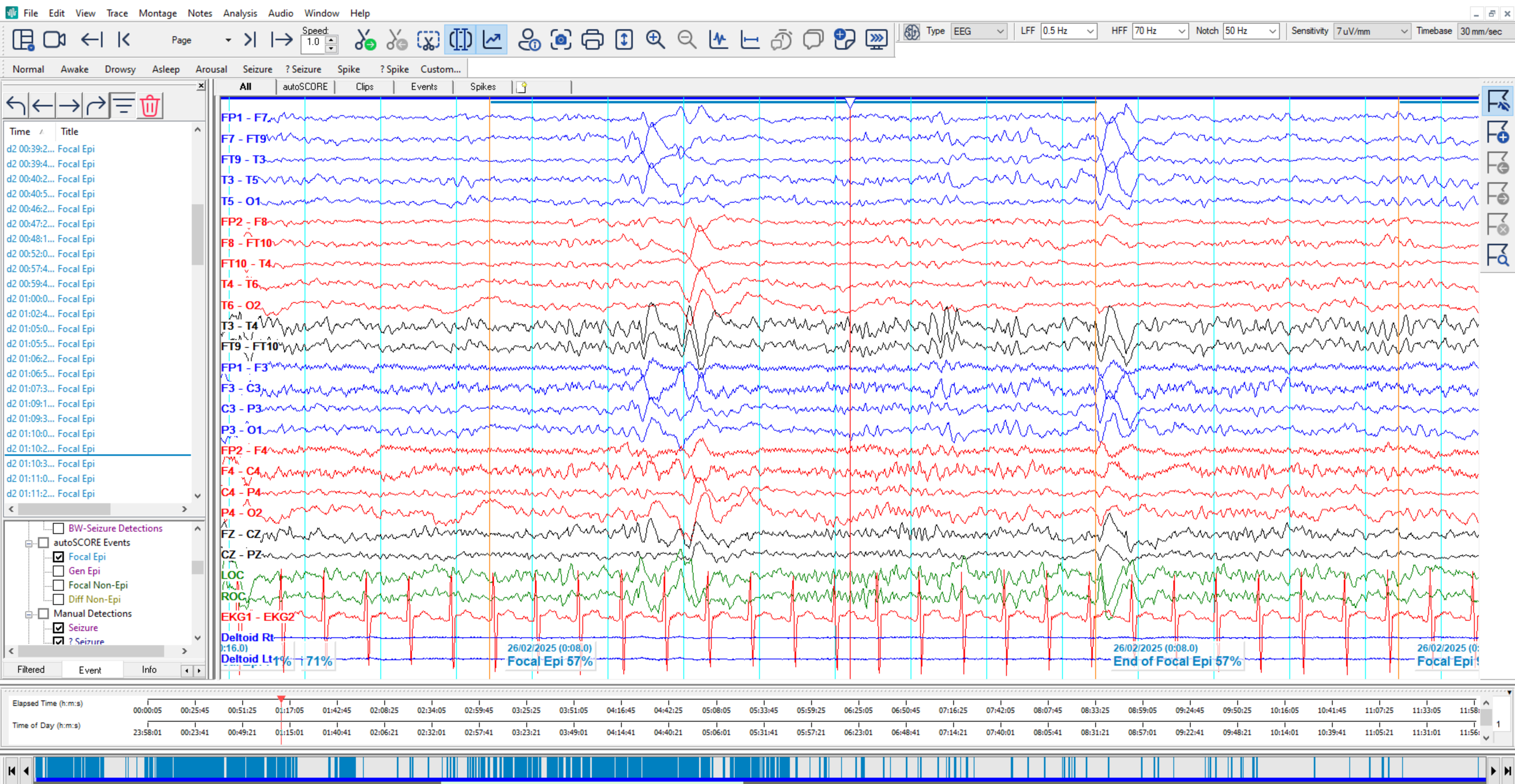
Table 3. Gwet AC1 Agreement Coefficients Between SCORE-AI and Clinical Assessment

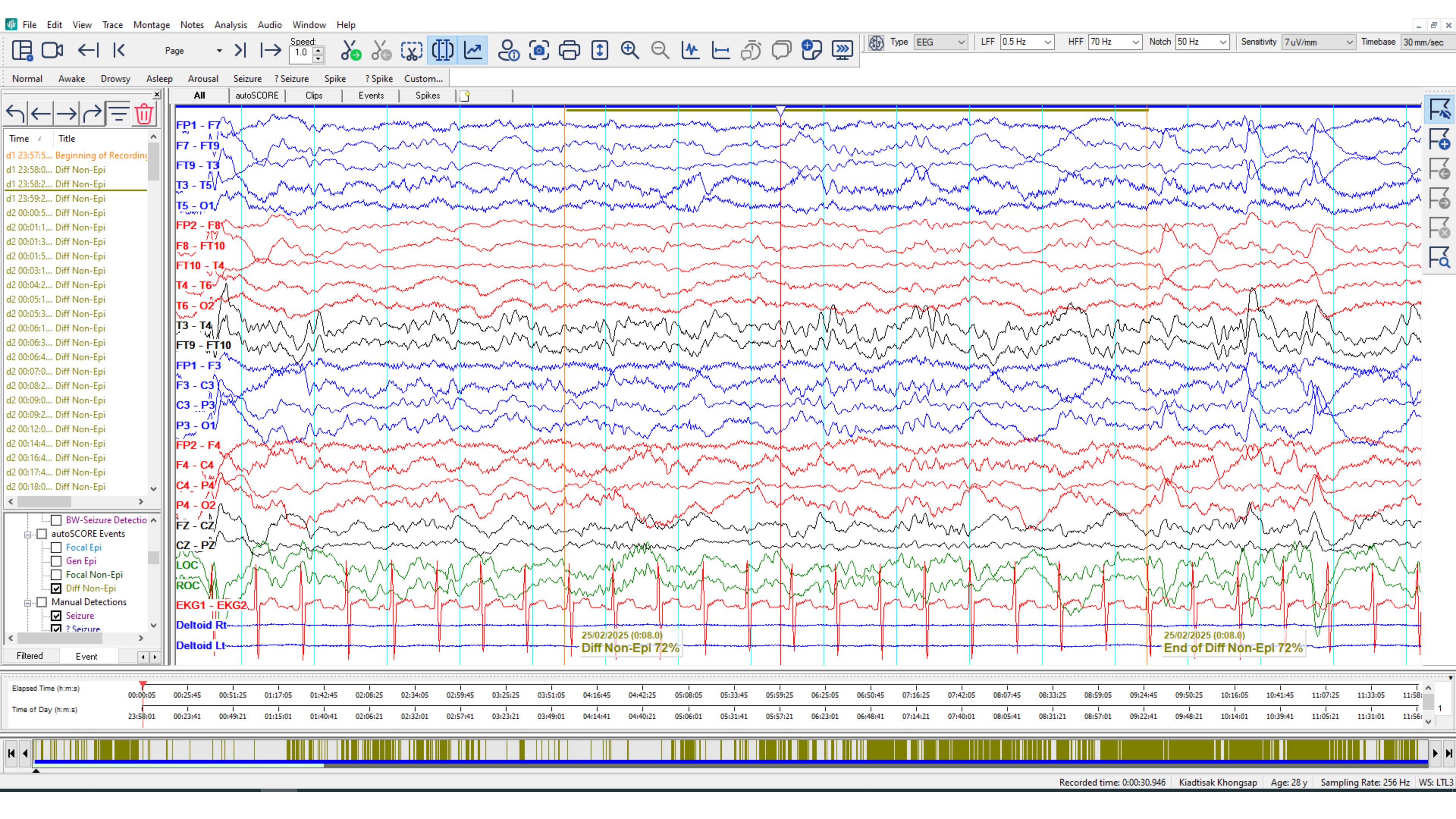
EEG recording category	Agreement between SCORE-AI and the clinical assessment of the EEGs	Difference between SCORE-AI-HE agreement and HE-HE agreement ^a
Normal	0.737 (0.723 to 0.750)	0.014 (−0.061 to 0.089)
Epileptiform-focal	0.871 (0.862 to 0.879) ^b	0.147 (0.067 to 0.228) ^b
Epileptiform-generalized	0.948 (0.943 to 0.953)	0.0471 (−0.001 to 0.095)
Nonepileptiform-diffuse	0.737 (0.723 to 0.750) ^b	0.106 (0.014 to 0.199) ^b
Nonepileptiform-focal	0.768 (0.756 to 0.780) ^b	0.181 (0.092 to 0.269) ^b
Exact match/multiple abnormalities	0.637 (0.627 to 0.647) ^b	0.140 (0.075 to 0.205) ^b

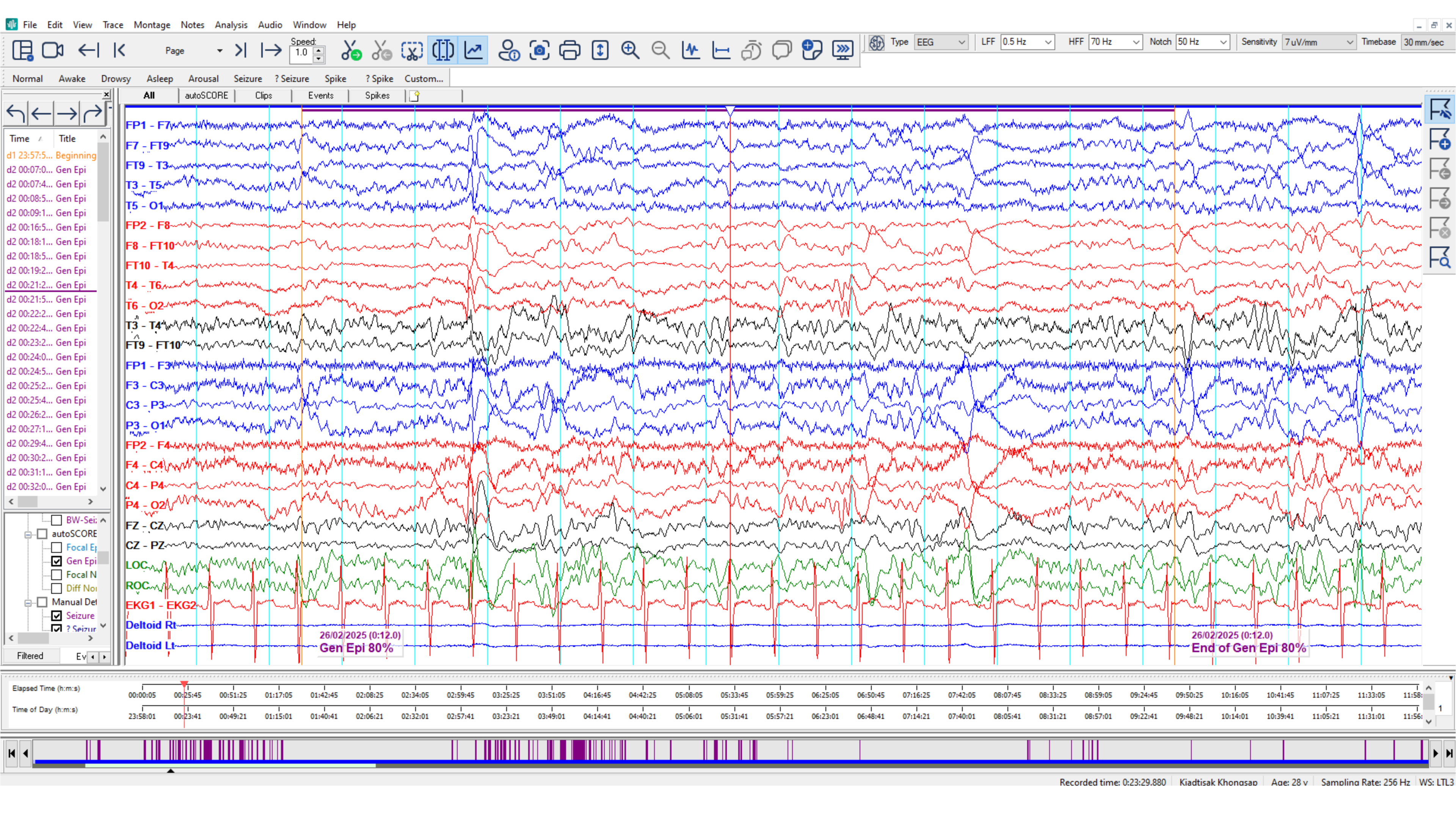
Abbreviations: EEG, electroencephalography; HE, human experts; SCORE-AI, Standardized Computer-based Organized Reporting of EEG–Artificial Intelligence.

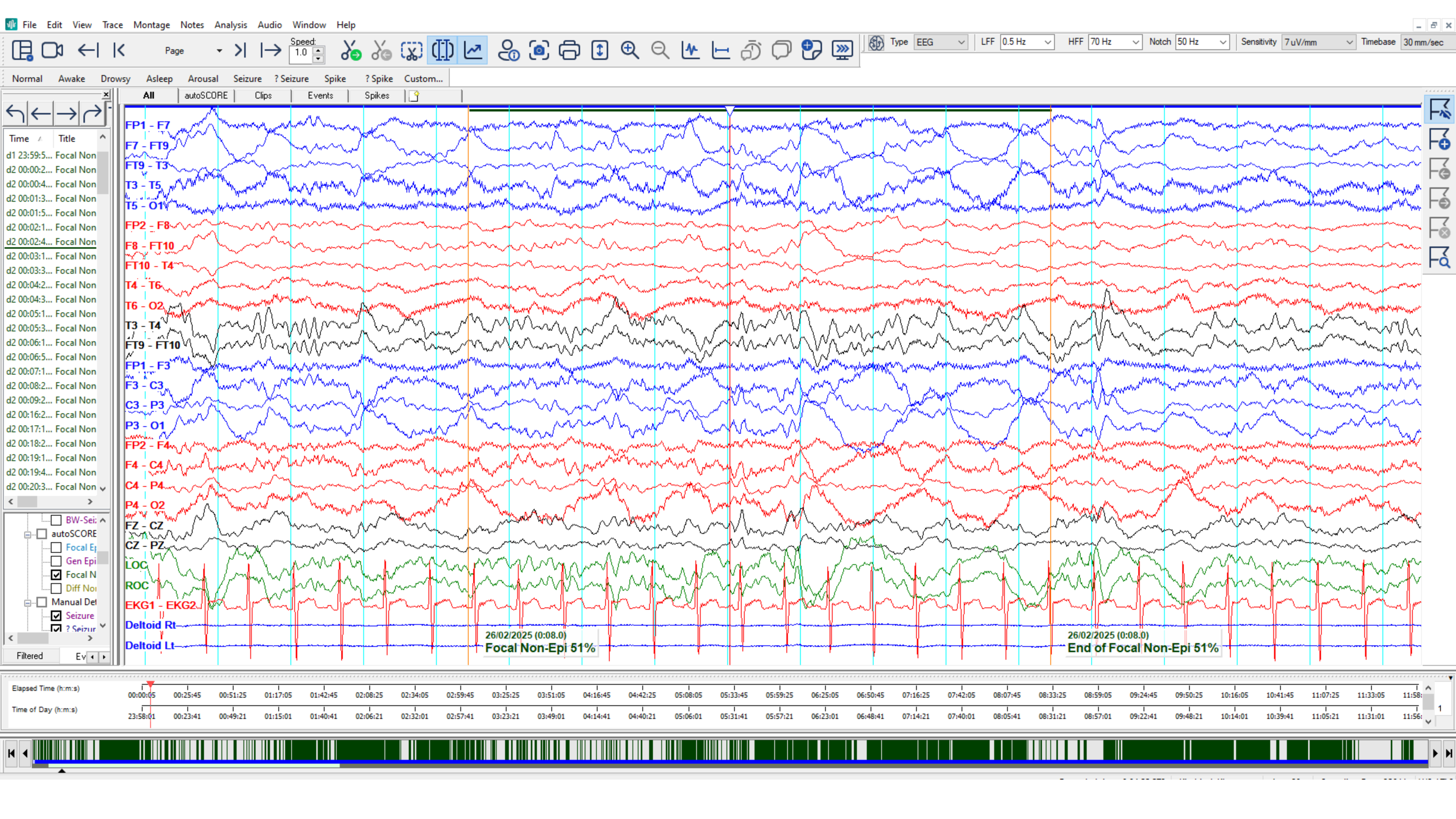
^a HE-HE agreement as detailed in Table 1.

^b Significant difference. Statistical comparisons were based on the 95% CIs. Significance means there was no overlap between the 95% CIs.







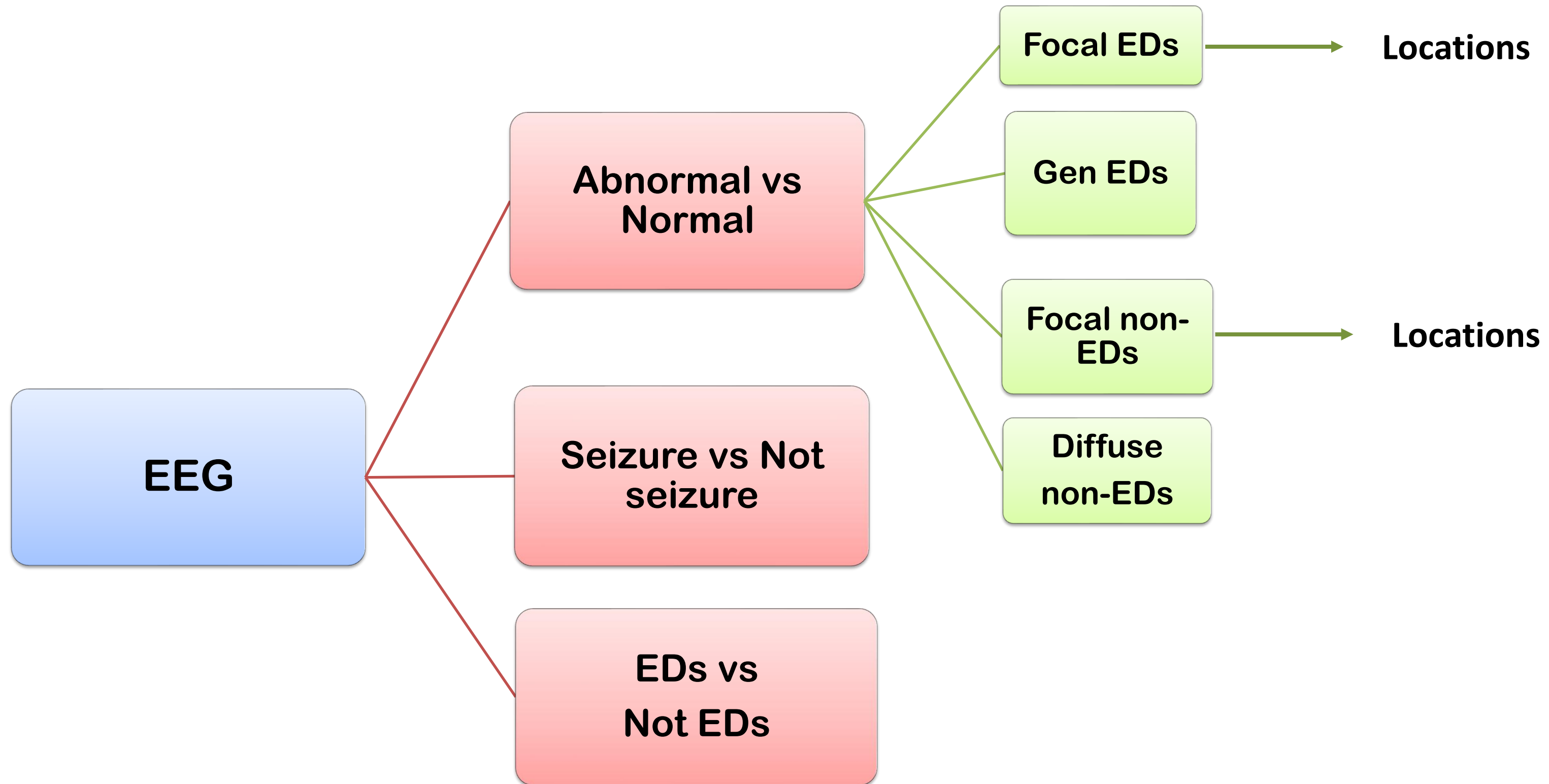


Promising for developing



(AutoScore®)

Ambulatory setting (Routine EEG at OPD)





Prediction score for determining seizure risk and the necessity of EEG monitoring



Table 1. 2HELPS2B

Risk Factor	Score	%					
		0	1	2	3	4	5+
Frequency >2 Hz ^a	1						
Independent sporadic epileptiform discharges	1						
LPD/BIPD/LRDA	1						
Plus features (superimposed rhythmic, fast, sharp) ^b	1						
Prior seizure ^c	1						
BIRD	2						
Total Score							
Predicted risk of seizure, ^d 2HELPS2B score		<5	12	27	50	73	88
Actual risk of seizure							
FS ^e		3	12	34	52	71	84
VAL ^f		4	15	34	55	75	93

"2HELPS2B"

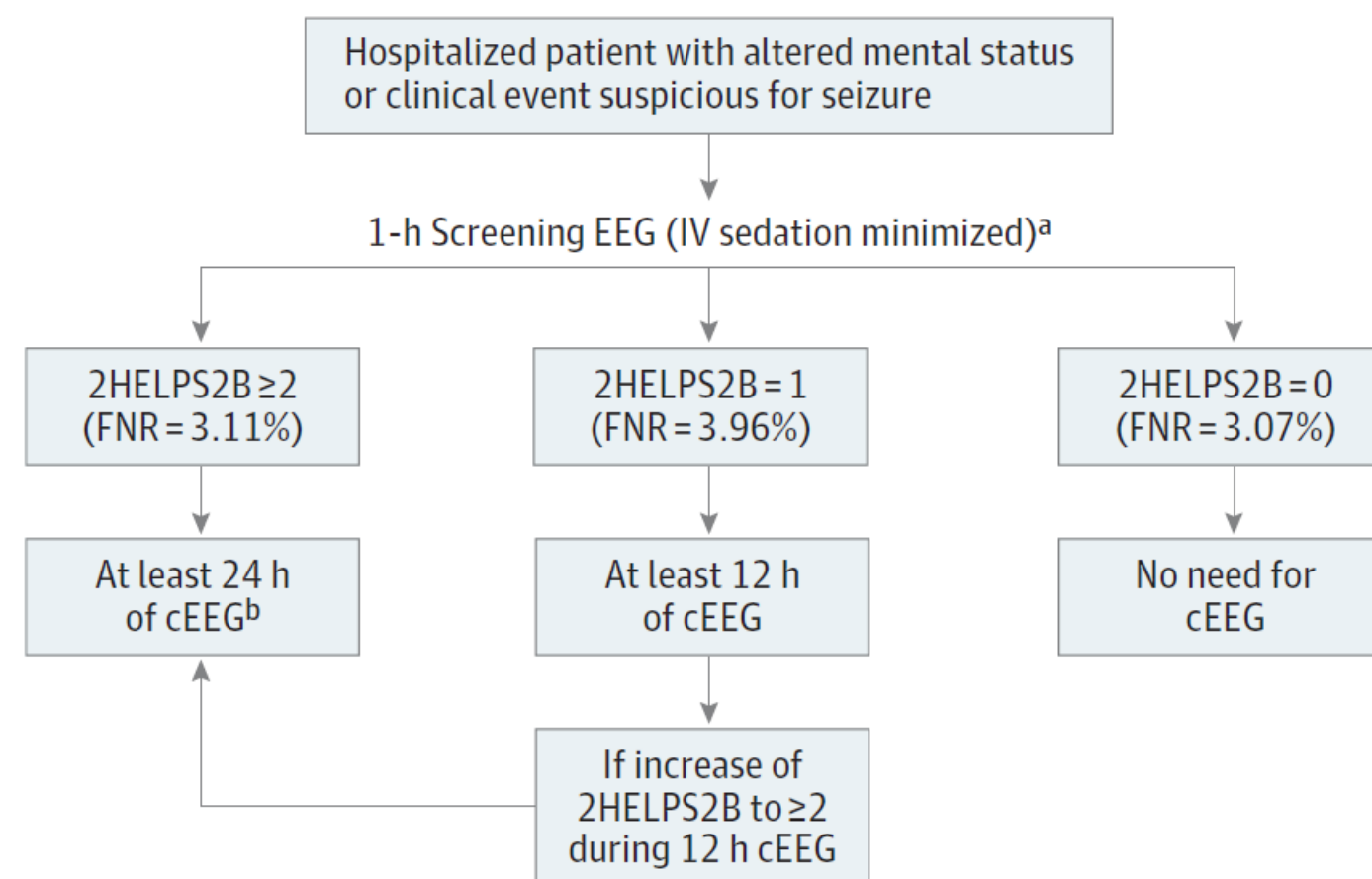
- ✓ > 2 Hz
- ✓ Epileptiform
- ✓ LPD
- ✓ Plus
- ✓ Seizure (prior)
- ✓ BIRD (2 points)

Recommend duration of EEG monitoring

Table 2. Seizure Risk Based on 1-Hour Screening EEG

Seizure Risk Group	No. (% of Cohort)	Overall Seizure Risk, %	False-Negative Rate, ^a %	Recommend Duration of EEG Monitoring
Low risk: ^b 2HELPS2B score = 0	594 (40)	3.1	3.1	1 h (Length of screening EEG)
Medium risk: ^c 2HELPS2B score = 1	597 (40)	12.0	4.0	12 h
High risk: ^d 2HELPS2B score, ≥ 2	310 (21)	26.6	3.1	At least 24 h

Figure 4. 2HELPS2B Clinical Algorithm



Struck AF et al., JAMA Neurol 2017
Struck AF et al., JAMA Neurol 2020

SPaRCNet

(“SPaRC” stands for Seizures, Periodic and Rhythmic pattern Continuum)

Development of Expert-Level Classification of Seizures and Rhythmic and Periodic Patterns During EEG Interpretation

Jin Jing, PhD,* Wendong Ge, PhD,* Shenda Hong, PhD, Marta Bento Fernandes, PhD, Zhen Lin, Chaoqi Yang, Sungtae An, Aaron F. Struck, MD, Aline Herlopian, MD, Ioannis Karakis, MD, PhD, MSc, Jonathan J. Halford, MD, Marcus C. Ng, MD, Emily L. Johnson, MD, Brian L. Appavu, MD, Rani A. Sarkis, MD, MSc, Gamaleldin Osman, MD, MS, Peter W. Kaplan, MBBS, FRCP, Monica B. Dhakar, MD, MS, Lakshman Arcot Jayagopal, MD, Zubeda Sheikh, MD, MS, Olga Taraschenko, MD, PhD, Sarah Schmitt, MD, Hiba A. Haider, MD, Jennifer A. Kim, MD, PhD, Christa B. Swisher, MD, Nicolas Gaspard, MD, PhD, Mackenzie C. Cervenka, MD, Andres A. Rodriguez Ruiz, MD, Jong Woo Lee, MD, PhD, Mohammad Tabaeizadeh, MD, Emily J. Gilmore, MD, Kristy Nordstrom, AS, Ji Yeoun Yoo, MD, Manisha G. Holmes, MD, Susan T. Herman, MD, Jennifer A. Williams, MB, BAO, Bch, FRCPI, Jay Pathmanathan, MD, PhD, Fábio A. Nascimento, MD, Ziwei Fan, MS, Samaneh Nasiri, PhD, Mouhsin M. Shafi, MD, PhD, Sydney S. Cash, MD, PhD, Daniel B. Hoch, MD, PhD, Andrew J. Cole, MD, Eric S. Rosenthal, MD, Sahar F. Zafar, MD, Jimeng Sun, PhD,† and M. Brandon Westover, MD, PhD†

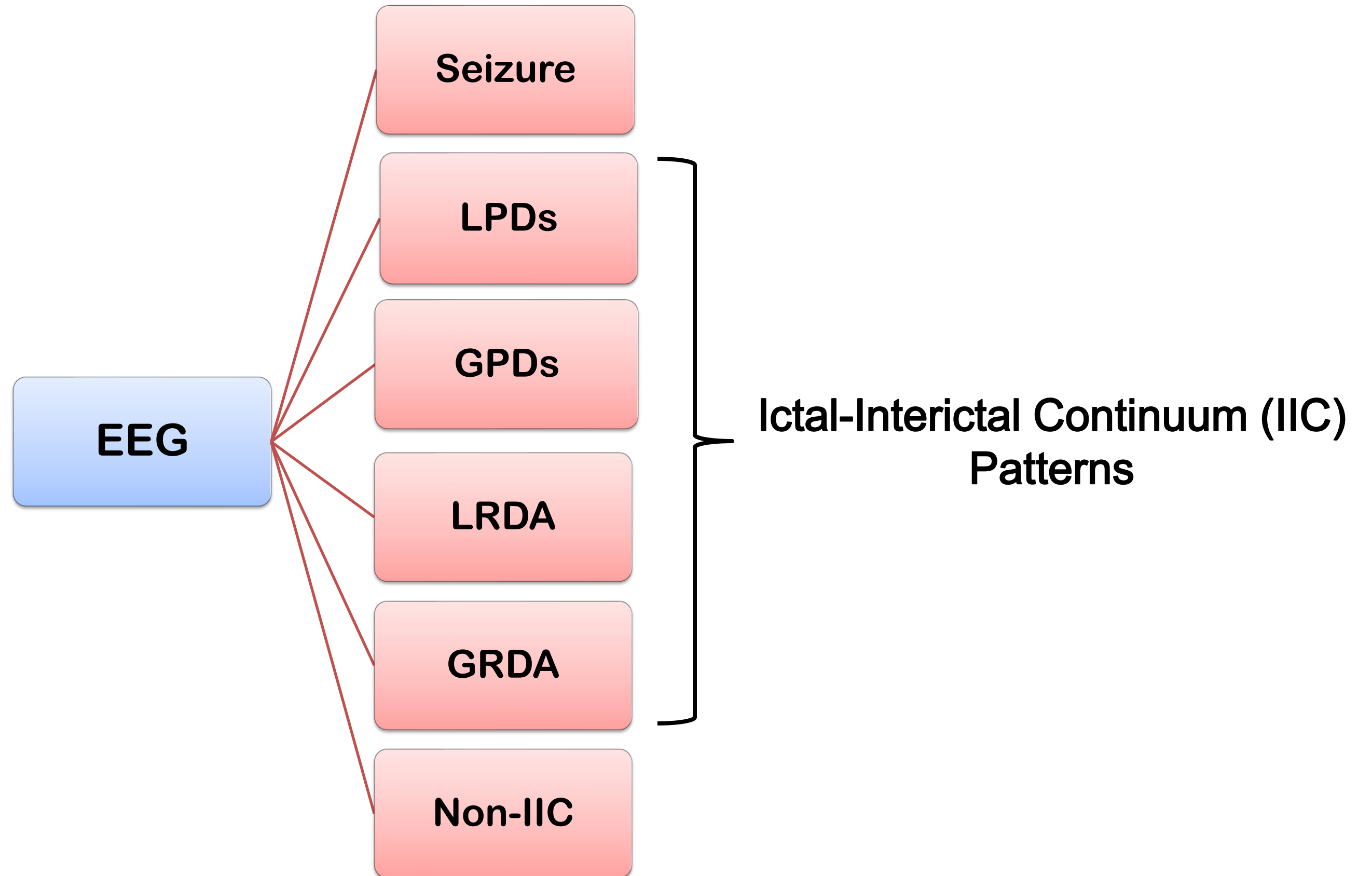
Correspondence

Dr. Westover
mwestover@
mgh.harvard.edu

Neurology® 2023;100:e1750-e1762. doi:10.1212/WNL.0000000000207127

(SPaRCNet)

Critically ill (Continuous EEG)

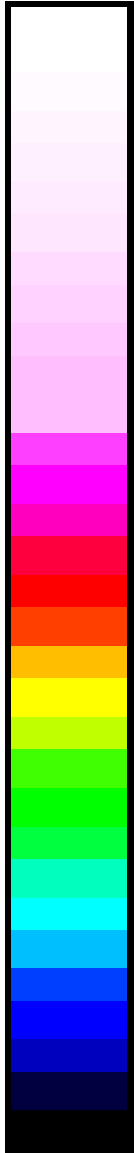


Promising for developing



Critical care

Dynamic changes
of cerebral functions



P...	T...	autoSCORE Conclusion	Focal Epi	Gen Epi	Focal Non-Epi	Diff Non-Epi
		<div><div></div>57 %</div>			<div><div></div>53 %</div>	<div><div></div>50 %</div>
		<div><div></div>97 %</div>	<div><div></div>91 %</div>	<div><div></div>67 %</div>	<div><div></div>67 %</div>	<div><div></div>80 %</div>
		<div><div></div>100 %</div>	<div><div></div>97 %</div>	<div><div></div>76 %</div>	<div><div></div>59 %</div>	<div><div></div>92 %</div>
		<div><div></div>99 %</div>	<div><div></div>96 %</div>	<div><div></div>60 %</div>	<div><div></div>76 %</div>	<div><div></div>69 %</div>

Remote EEG monitoring



Rationales

If the hypothesis is true

Early NCS/NCSE detection by EEG



Lead to early and appropriate treatment
for NCS/NCSE



Better functional outcomes

EEG is still not feasible to be used in
everywhere since it essentially
requires EEG specialists to interpret
the findings

Solutions

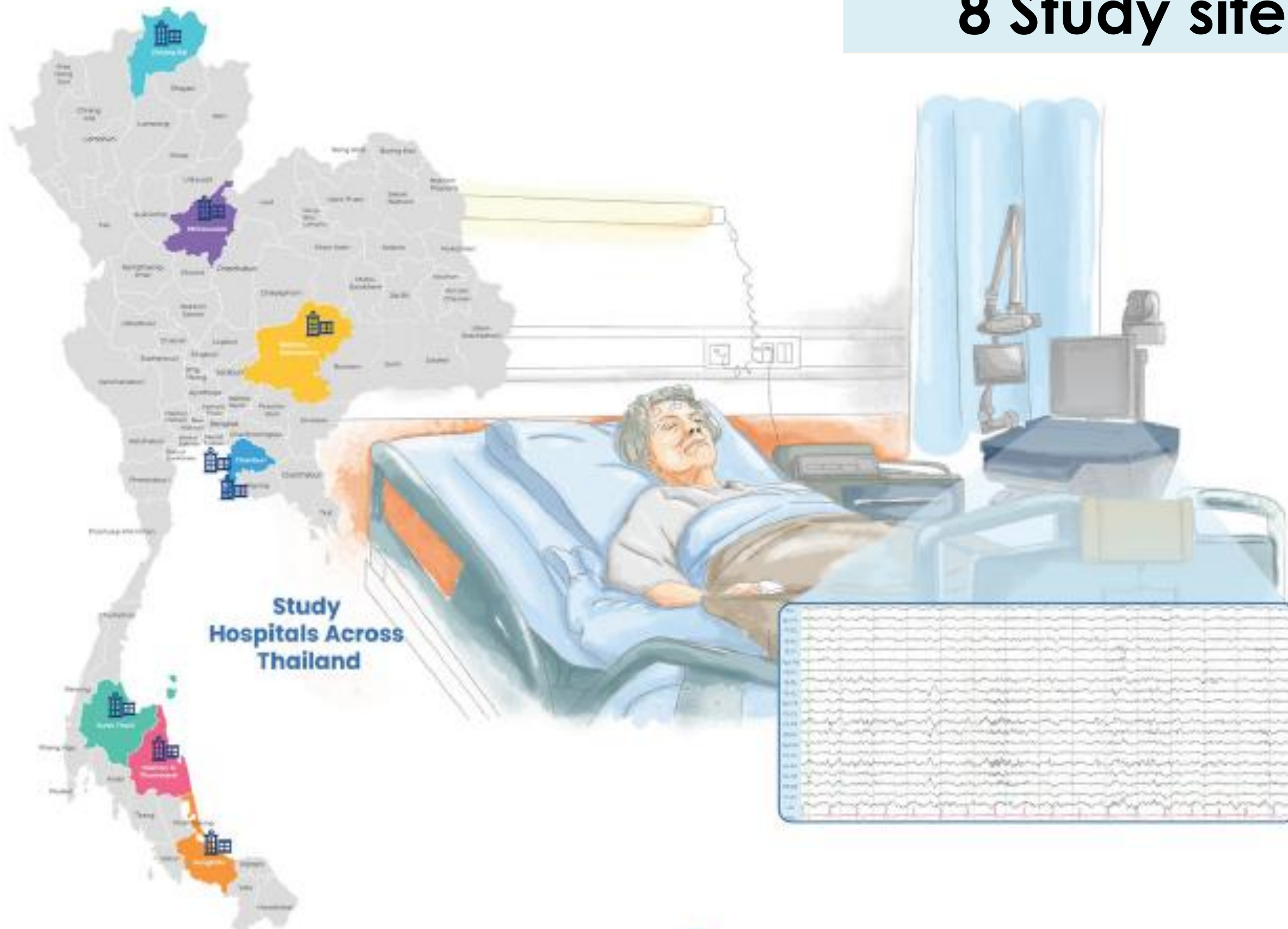
Tele-EEG

EEG
help detect seizures and
guide for appropriate
management



Telemedicine (TM)
Delivery of care to distant
hospitals by specialists

8 Study sites





Functional outcome and mortality assessment



30 MO RECRUITMENT

1. Age ≥ 15 years
2. Suffers from at least one of the following conditions; recent clinical seizure/SE without return to baseline, severely depressed LOC from any cause, intracranial haemorrhages, clinically suspicious of NCS/NCSE, CNS infection with altered mental status
3. Without following conditions; post cardiac arrest, advanced cancer, AIDS, alcohol intoxication, poor functional outcome (mRS 4-6), extensive surgical wounds

Signed consent obtained, EEG technician and ICU/ward beds are available, then perform central randomization

Start EEG recording
< 24 h after randomization

Tele-cEEG
(N = 128)

Tele-cEEG
24-72 h

Switch
(N = 43)

Tele-rEEG
30 min

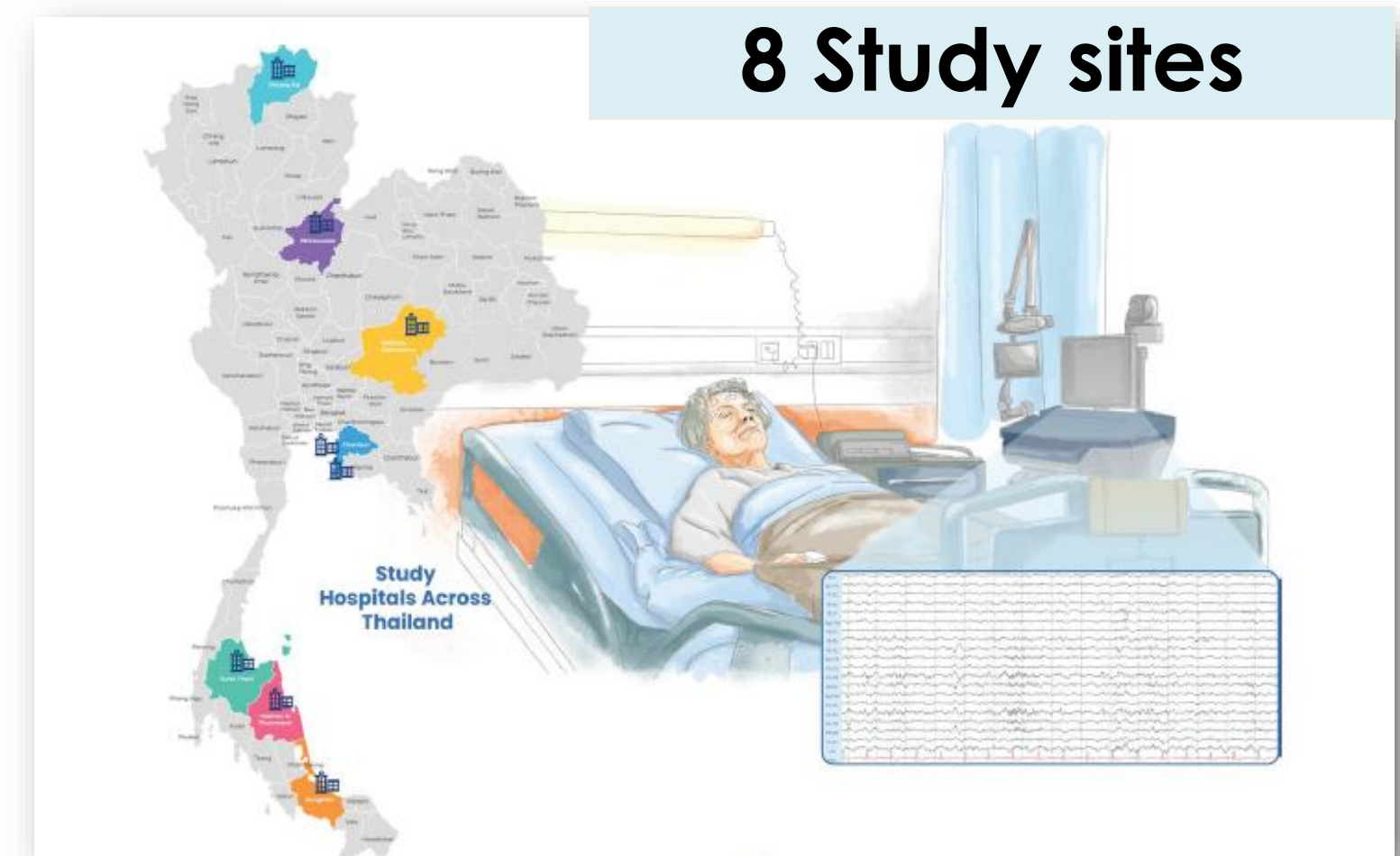
Tele-rEEG
(N = 126)

8 patients:
recording > 72 h

Primary outcome

- ✓ Seizure detection rate
- ✓ Mortality
- ✓ Functional outcome (mRS)

8 Study sites



Limotai C et al., Critical Care 2025

RESEARCH

Open Access



Efficacy of delivery of care with Tele-continuous EEG in critically ill patients: a multicenter randomized controlled trial (Tele-cRCT study) study

Chusak Limotai^{1,2,3}, Suda Jirasakuldej^{2,3}, Sattawut Wongwiangiunt⁴, Tipakorn Tumnark⁵, Piradee Suwanpakdee⁶, Kwuanrat Wangponpattanasiri⁷, Piyanuch Rakchue⁸, Chaiwiwat Tungkasereerak⁹, Polchai Pleumpanupatand¹⁰, Phopsuk Tansuhaj¹¹, Phattarawin Ekkachon¹², Songchai Kittipanprayoon¹³, Apiwoot Kerddonfag¹⁴, Thippamas Pobsuk¹⁵, Anuchate Pattanateepapon¹, Kammant Phanthumchinda³, Nijasri C. Suwanwela³, Iyavut Thaipisuttikul³, Kanokwan Boonyapisit⁴, Atiporn Ingsathit¹, Oraluck Pattanapratchee¹, John Attia¹⁶, Gareth J. McKay¹⁷, Andrea O. Rossetti¹⁸, Ammarin Thakkinstantian^{1*}, Chutima Rukrun¹, Patcharapun Kangsanant¹, Jeerawan Mokkaew¹, Nittaya Phayaph¹, Supak Pukpraman¹, Warangkana Ritthathon¹, Youwarat Jarungjitapinan¹, Jintana Pinpradab¹, Netphit Khamhoi¹, Mayuree Nookaew¹, Patchareeporn Chauywang¹, Pichai Rojdmapiyakorn¹, Paworamon Sribussara¹, Wasunon Tinroongroj¹, Wisan Teeratanthakorn¹, Tabtim Chongsuvivatwong¹, Watchara Viratyaporn¹, Witoon Jantararotai¹, Komkrit Panyawattanakit¹, Nopparat Rujirarongrueng¹, Pornnapat Damthong¹, Pattama Udom¹, Molviba Siensuwan¹, Phatcharamai Phonprasori¹, Karnpidcha Wanmuang¹.

Limotai C et al.,
Critical Care 2025

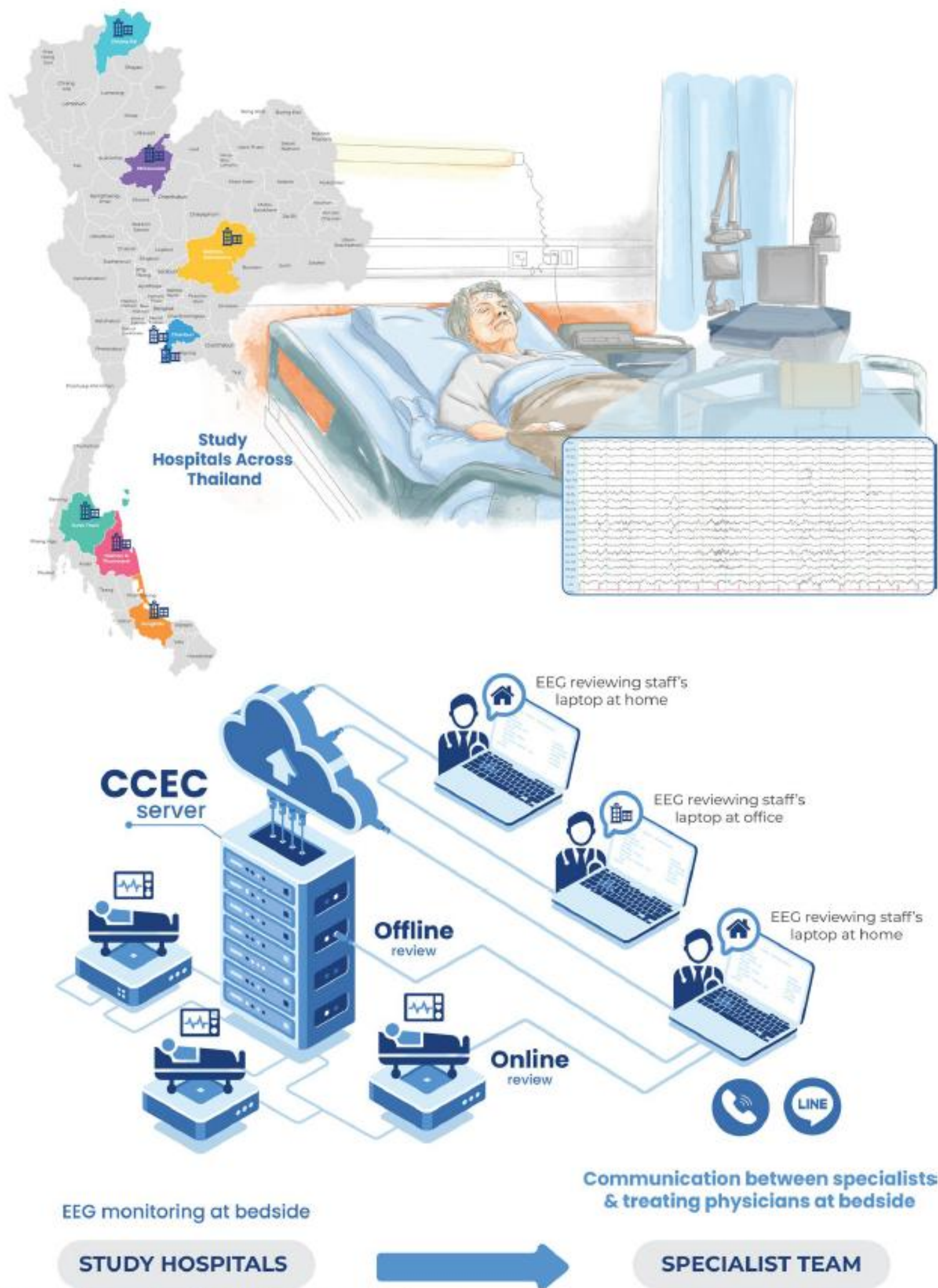
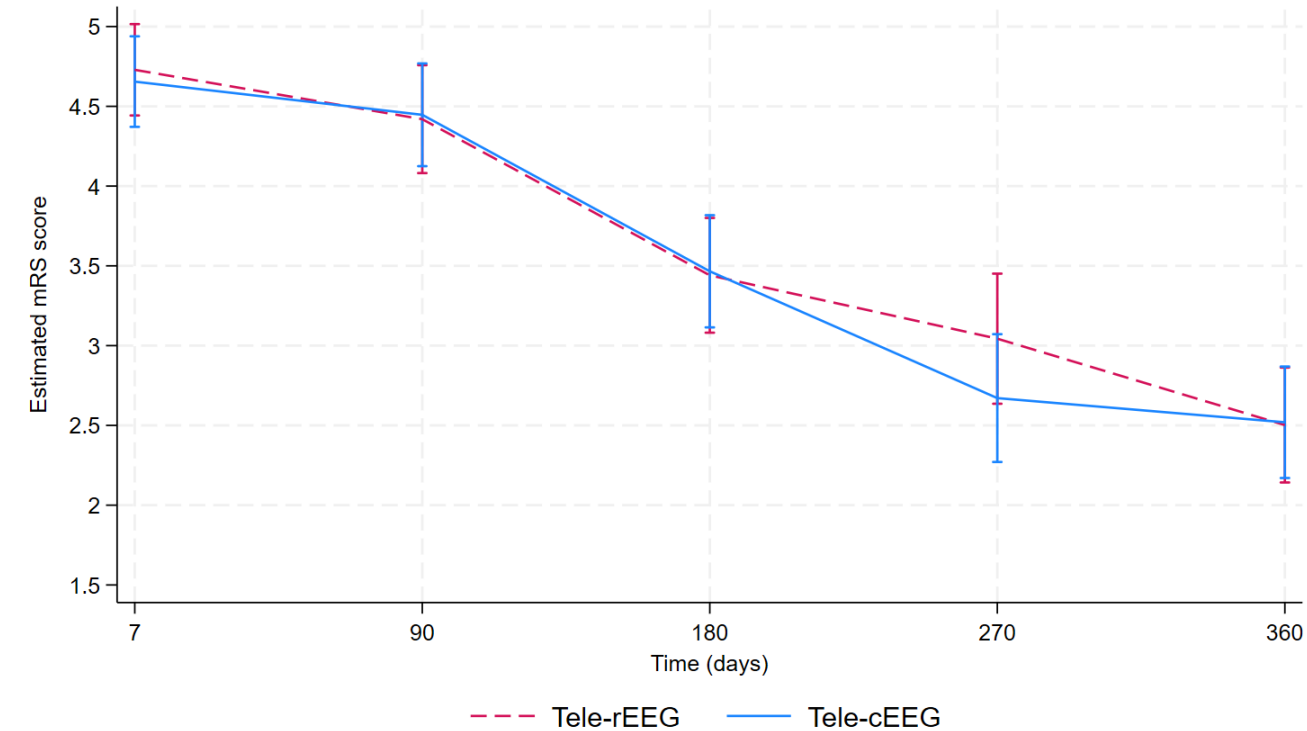
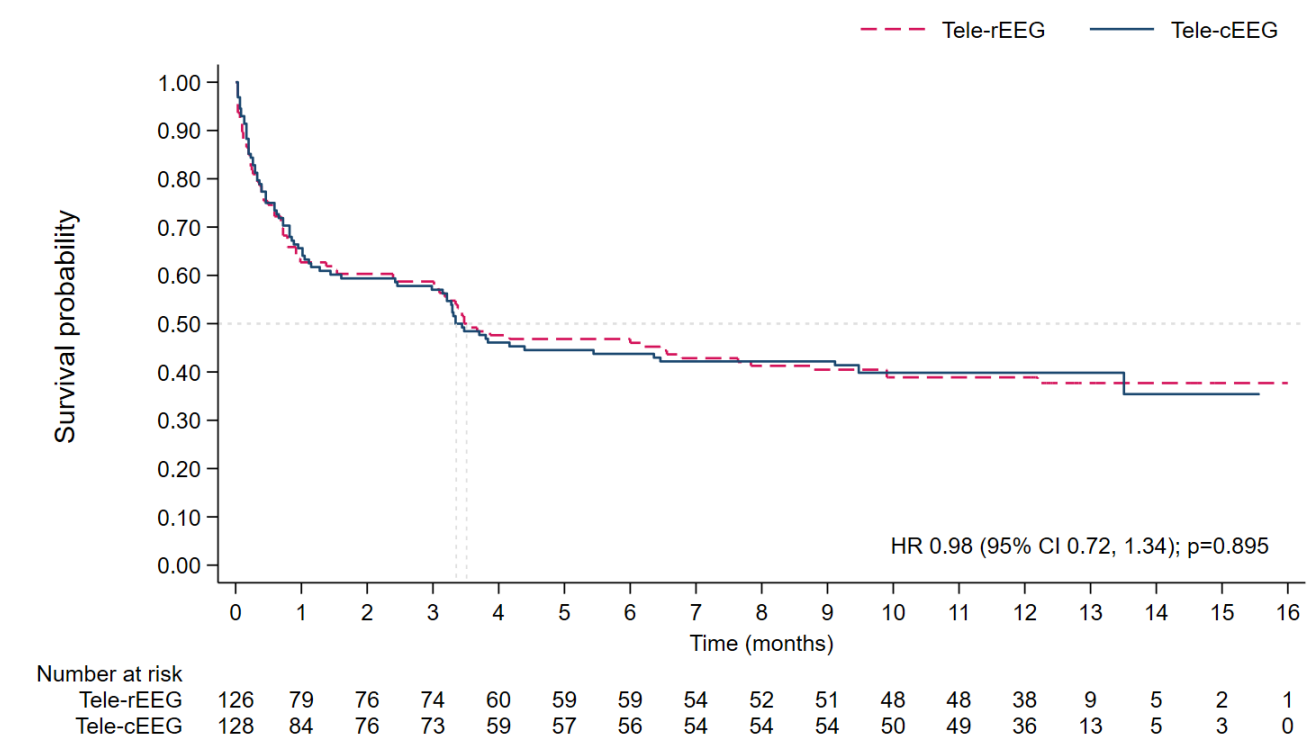


Fig. 1 Pictorial demonstration of the Tele-EEG and communication system



**Limotai C et al.,
Critical Care 2025**



Conclusions:

- While Tele-cEEG may help detect NCS/NCSE, this study had limited power to detect its efficacy in reducing mortality or improving functional outcomes.
- In limited-resource settings**, Tele-rEEG approximating 30 min or longer offers a feasible and potentially valuable initial screening tool for critically ill patients at-risk of seizures. However, where Tele-cEEG is readily available, it remains the recommended approach.

Wearable devices



NEJM 2024

The NEW ENGLAND JOURNAL of MEDICINE

REVIEW ARTICLE

WEARABLE DIGITAL HEALTH TECHNOLOGIES IN MEDICINE

Stephen H. Friend, M.D., Ph.D., *Guest Editor*, Geoffrey S. Ginsburg, M.D., Ph.D., *Guest Editor*,
Rosalind W. Picard, Sc.D., *Guest Editor*, and Jeffrey M. Drazen, M.D., *Editor*

Wearable Digital Health Technology for Epilepsy

Elizabeth Donner, M.D., Orrin Devinsky, M.D., and Daniel Friedman, M.D.

The NEW ENGLAND JOURNAL of MEDICINE

REVIEW ARTICLE

WEARABLE DIGITAL HEALTH TECHNOLOGIES IN MEDICINE

Stephen H. Friend, M.D., Ph.D., *Guest Editor*, Geoffrey S. Ginsburg, M.D., Ph.D., *Guest Editor*,
Rosalind W. Picard, Sc.D., *Guest Editor*, and Jeffrey M. Drazen, M.D., *Editor*

Key Issues as Wearable Digital Health Technologies Enter Clinical Care

Geoffrey S. Ginsburg, M.D., Ph.D., Rosalind W. Picard, Sc.D., and
Stephen H. Friend, M.D., Ph.D.

Wearable Digital Health Technologies (DHTs)



Physiological signals assessed by peripheral sensors

*Sensors currently used in commercial seizure wearables

- EEG*
- Audio
- Heart rate*
- Muscle contraction*
- Movement*
- Breathing and oxygen saturation
- Skin conductance*



Wearable device options

- Armband
- Wristwatch

Multiple sensor types in a wearable device can detect a possible seizure, which wirelessly notifies a paired smartphone

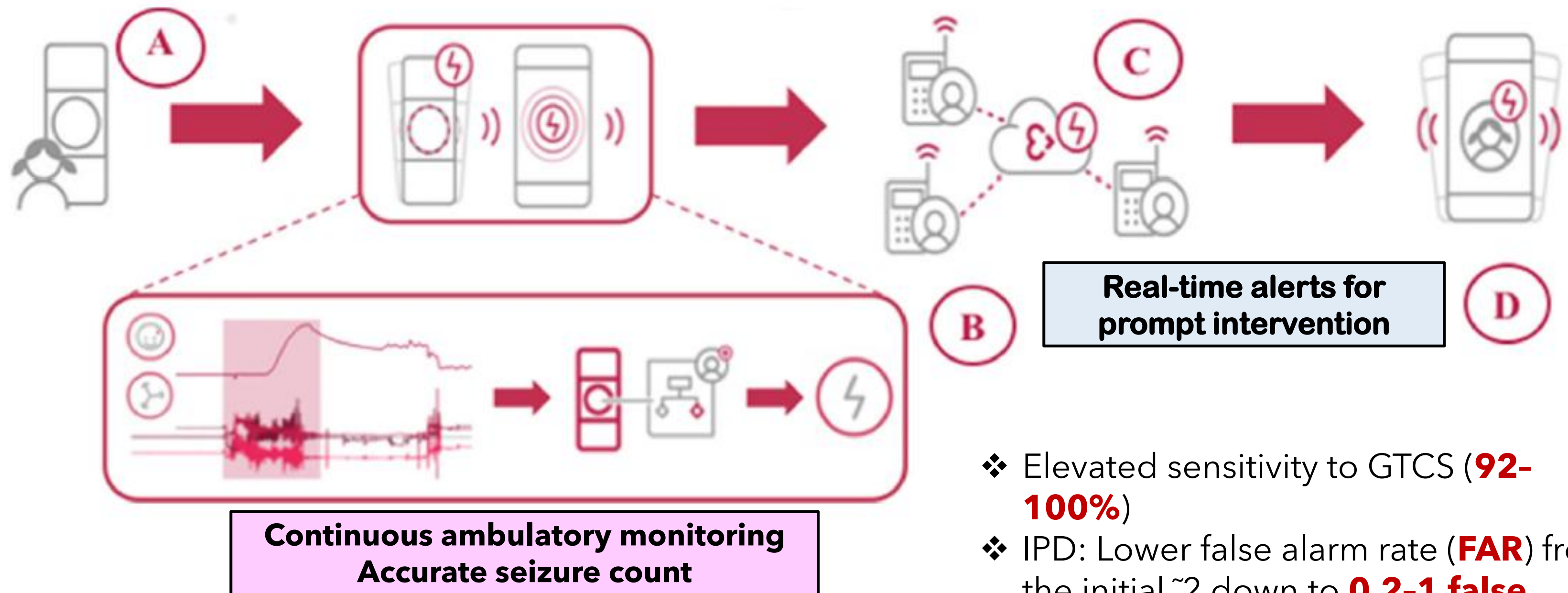
Embrace® received clearance by the US FDA in 2018 (Class II) for **GTCS** monitoring during periods of rest for adults and children aged 6 and up



Automatically alerts your loved ones when you have a seizure



A “machine learning algorithm” able to recognize ACC and EDA signatures of **GTCS-like events**



- ❖ Elevated sensitivity to GTCS (**92-100%**)
- ❖ IPD: Lower false alarm rate (**FAR**) from the initial ~2 down to **0.2-1 false alarms/day**
- ❖ OPD: initial ~6 down to < 0.5 false alarms/day

Wrist-worn smartwatch and predictive models for seizures

Waroth Pipatpratarnporn¹ | Wichuta Muangthong¹ | Suda Jirasakuldej² |
Chusak Limotai^{1,2} 










PIPATPRATARNPORN ET AL.

TABLE 4 Univariate and multivariate analysis using generalized estimating equations and best-fit predictive models for each seizure type.

Seizure types	Univariate analysis		Multivariate analysis		Best-fit model and cut-points
		<i>p</i>		<i>p</i>	
Over all seizure types	HR	<.001	HR	<.001	$y = -4.507 + .046$ (HR) + 24.570 (ACC) Cut-point <i>y</i> value for overall seizures $\geq -.35$
	ACC	.001	ACC	.008	
	EDA	.092			
	TEMP	.651			
BTCs	HR	.017	ACC	.005	$y = -6.772 + 33.066$ (ACC) + .040 (HR) Cut-point <i>y</i> value for BTCs ≥ -2.47
	ACC	.001	HR	.133 ^a	
	EDA	.173			
	TEMP	.466			
Non-BTCs	HR	.001	ACC	.003	$y = -4.344 + 28.554$ (ACC) + .037 (HR) Cut-point <i>y</i> value for non-BTCs $\geq -.87$
	ACC	<.001	HR	.025	
	EDA	.213			
	TEMP	.785			
Isolated auras	HR	<.001	HR	<.001	$y = -7.838 + .0691$ (HR) Cut-point <i>y</i> value for isolated auras ≥ -1.69
	ACC	.916			
	EDA	.883			
	TEMP	.789			

Wearable devices for seizure detection: Practical experiences and recommendations from the Wearables for Epilepsy And Research (WEAR) International Study Group

Elisa Bruno^{1*}  | Sebastian Böttcher^{2,3*}  | Pedro F. Viana^{1,4}  |
Marta Amengual-Gual⁵  | Boney Joseph⁶  | Nino Epitashvili^{2*}  |
Matthias Dümpelmann^{2*}  | Martin Glasstetter²  | Andrea Biondi^{1*}  |
Kristof Van Laerhoven³  | Tobias Loddenkemper⁵  | Mark P. Richardson^{1*}  |
Andreas Schulze-Bonhage^{2*}  | Benjamin H. Brinkmann⁶ 

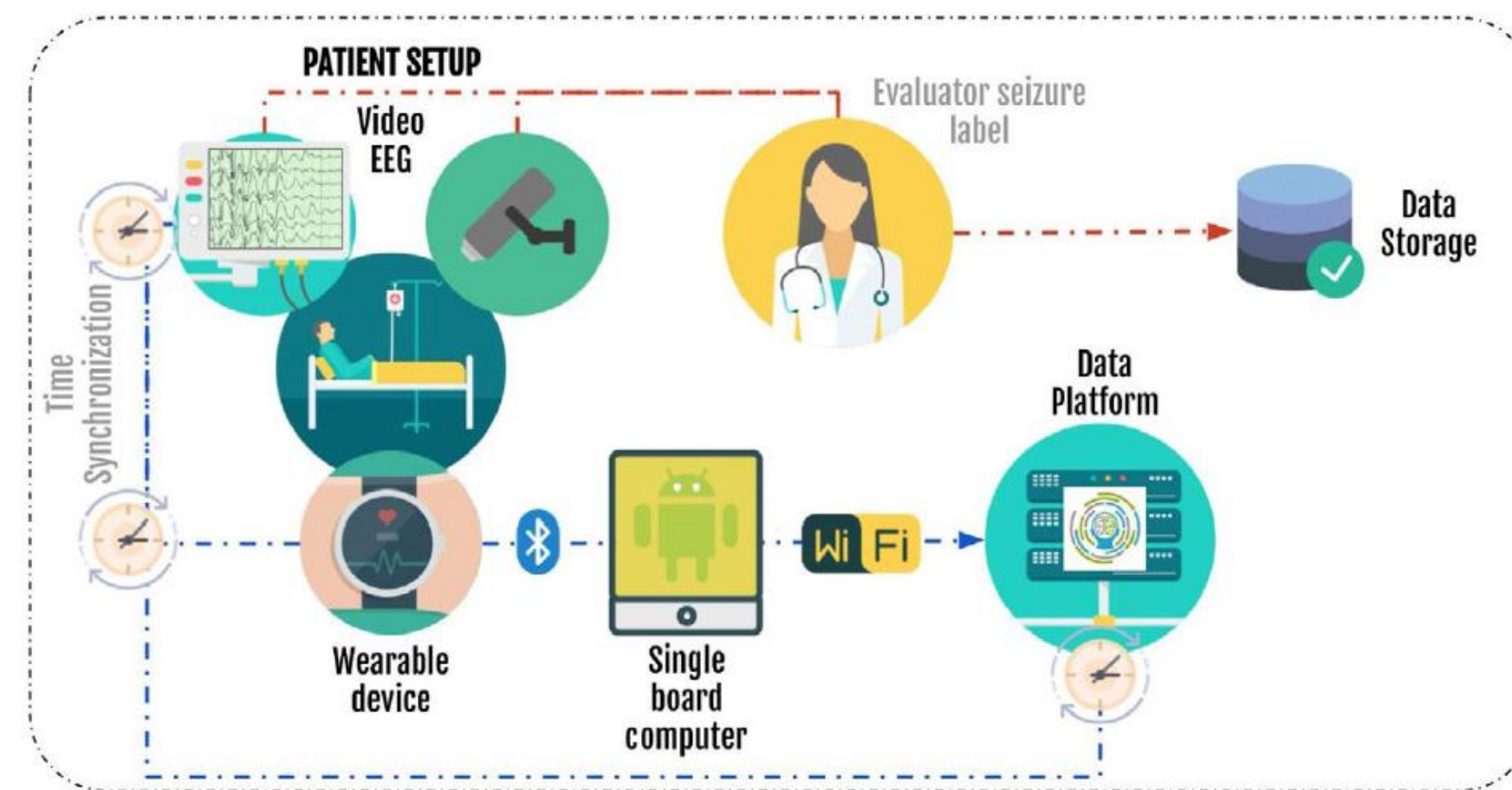
The Wearables for Epilepsy And Research (WEAR)

International Study Group identified a set of methodology standards to guide research on wearable devices for seizure detection

We formed an international consortium of experts from clinical research, engineering, computer science, and data analytics at the beginning of 2020

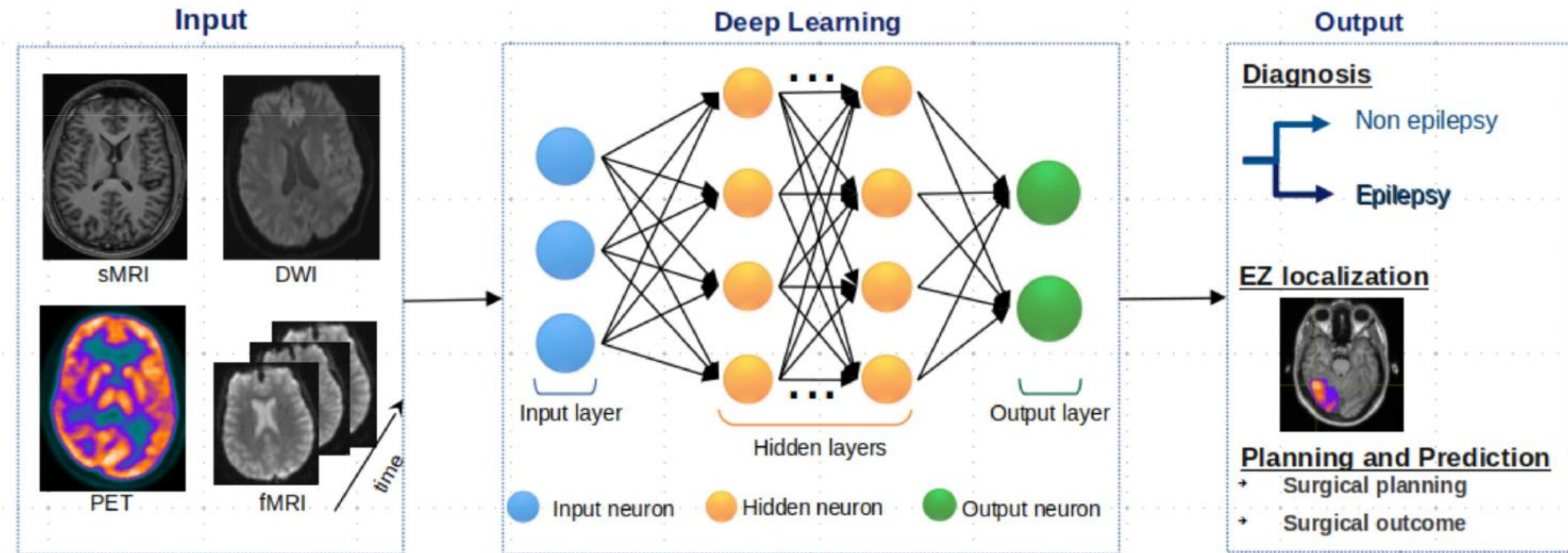
Introducing a **framework of methodology standards** promotes optimal, accurate, and consistent data collection. It also guarantees that studies are **generalizable and comparable, and that results can be replicated, validated, and shared.**

Setup of the technical environment for in-hospital studies on wearable devices for seizure detection (setting epilepsy monitoring unit)





Tools for Epilepsy Surgery



Structural MRI: Diagnosis of FCD, TLE, HS, JME (connectivity)

Table 1

Published studies for diagnosis/lateralization of epilepsy using DL models based on structural neuroimaging.

Modality	Ref.	ANN Input	DL model	Population and validation scheme	Performance	Ground truth
sMRI	[44]	T1 and FLAIR-based features	FFNN	20 FCD and 28 controls (70:15:15%) External validation: 10 FCD (negative MRI)	ACC _D ~ 91%	Pathologically confirmed cases
	[40]	T2	CNN	100 left HS-TLE, 60 right HS-TLE, 160 controls (5-fold cross-validation) External validation: 25 left HS-TLE, 25 right HS-TLE, 252 controls	ACC _D ~ 89% ACC _L ~ 91%	66% of patients pathologically confirmed
	[42]	FLAIR	CNN based transfer learning	85 HS-TLE, 56 healthy subjects, 5-fold cross-validation (80:20%)	ACC _D ~ 87%	Pathologically confirmed cases
DWI	[43]	Structural connectivity	learning	33 JME, 30 healthy subjects (80:20%)	ACC _D ~ 92%	No reported
SMRI	[41]	T1-based features	FCN	187 left TLE, 149 right TLE, 631 controls (80:20%) 10-fold cross-validation	ACC _D ~ 73% ACC _L ~ 77%	Standard of care assessment batteries
vs DWI		DWI-based features		482 left TLE, 381 right TLE, 976 controls (80:20%) 10-fold cross-validation	ACC _D ~ 74% ACC _L ~ 66%	

TLE: Temporal lobe epilepsy. HS: hippocampal sclerosis. FCD: Focal cortical dysplasia. JME: juvenile myoclonic epilepsy. sMRI: structural Magnetic Resonance Imaging. DWI: Diffusion-Weighted Imaging. FLAIR: Fluid attenuated inversion recovery. CNN: Convolutional Neural Network. FFNN: Feed-Forward Neural Network. FCN: Fully Connected Neural Network. ACCD: Accuracy in diagnosis. ACCL: Accuracy in lateralization.

Functional Imaging: Diagnosis and Lateralization of TLE

Table 2

Published studies for diagnosis/lateralization of epilepsy using DL models based on functional neuroimaging.

Modality	Ref.	ANN Input	DL model	Population and validation scheme	Performance	Ground truth
fMRI	[45]	Temporal latency	CNN	63 pediatric patients, 259 controls (60:20:20%)	ACC _D ~ 74%	Standard of care assessment batteries at each site
	[46]	FC features	FFNN	46 pediatric patients leave-one-out cross validation	ACC _L ~ 89%	
	[47]	BOLD time series	CNN	2132 controls (80:20%) External validation: 20 left TLE, 12 right TLE	ACC _L ~ 90%	Video-EEG monitoring
[¹⁸ F]FDG-PET	[48]	ROI metabolism metrics	FFNN	39 left TLE, 34 right TLE, 32 controls Leave-one-out cross-validation	ACC _D ~ 76% ACC _L ~ 89%	Not reported
	[49]		CNN based Transfer learning	7 patients, 8 controls	ACC _D ~ 98%	

TLE: Temporal lobe epilepsy. fMRI: functional Magnetic Resonance Imaging. PET: Positron Emission Tomography. ROI: Region of interest. CNN: Convolutional Neural Network. FFNN: Feed-Forward Neural Network. ACCD: Accuracy in diagnostic. ACCL: Accuracy in lateralization. FC: functional connectivity. BOLD: Blood oxygenation level dependent.

Surgical Planning and Outcomes

Table 5

Published studies for surgical planning and prediction of surgical outcomes using DL techniques.

Aim	Ref.	ANN Input	Population and validation scheme	DL model	Performance
Detecting eloquent white matter pathways of interest	[67]	DWI streamlines coordinates	70 healthy children (80:20%) External testing: 70 children with focal epilepsy	CNN	ACC ~ 73–100%
	[68]		89 children with focal epilepsy (45:18:37%)		ACC ~ 86–100%
Predict expressive (e) and receptive (r) language scores	[69]	Structural connectivity	37 children with focal epilepsy (70:30%) leave-one-out cross validation		Re/r = 0.82/ 0.75 MAEe/r = 7.8/ 7.4
	[70]		31 children with focal epilepsy (65:35%) 10-fold cross validation	CNN+RN	Re/r = 1/0.99 MAEe/r = 0.3/ 0.3
Predict language scores and postoperative seizure	[71]		51 children with focal epilepsy (80:20%) 5-fold cross-validation		MAEe/r = 0.9/ 0.3 P ~ 89–93%
Predict postoperative seizure freedom	[74]		50 TLE (5-fold cross validation)	ANN	P ~ 88%
	[75]	Graph-theory measures	121 TLE (72:28%) External testing: 47 TLE	FFNN	AUC = 0.88

TLE: Temporal lobe epilepsy. DWI: Diffusion-Weighted Imaging. CNN: Convolutional Neural Network. RN: Relational Network. R: Pearson correlation coefficient. MAE: mean absolute error. P: precision. FFNN: Feed-Forward Neural Network. AUC: Area Under Curve.

●●● **Challenges implementing in practice**

- **Small sample size**
- **Very few studies performed external validation**



Large Language Models (ChatGPT)

AI Chatbot

FUNDAMENTAL:

Chatbots use large databases to train computational models **to seed the first word of a response to a prompt**, then predict the best or most likely word, or to stop the response

Training with large databases:

There are over 170,000 words in the English language; therefore, for a machine to interpret just two-word phrases, the algorithm likely needs to be trained on over **30-trillion word-pairs** (170,000 squared)

Limitations:

Likely rely on masses of publicly facing information and probably **exclude information that requires additional levels of access** (e.g., textbooks, subscription-based peer-reviewed literature requiring journal subscriptions)

Kerr WT & McFarlane KN; Current Neurology and Neuroscience Reports (2023)





Hallucinations

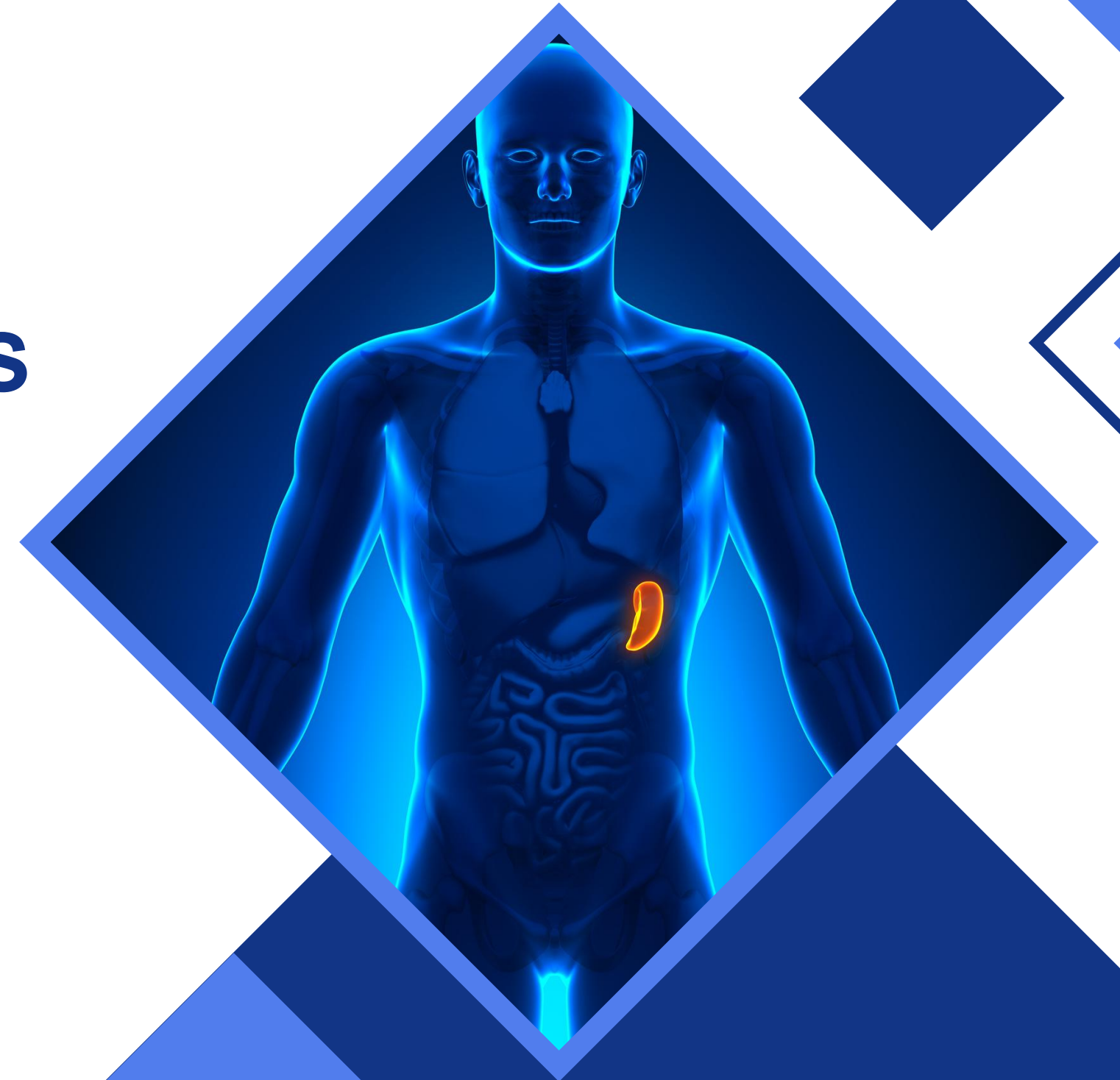
What's it?

AI responses that appear incorrect or nonsensical

Examples:

1. ChatGPT was asked for the surgical options to treat **bilateral temporal epilepsy**, it suggested **bilateral temporal lobectomy** but noted that the adverse effects of that operation may outweigh the benefit
2. When AI chatbots provided citations, they may cite manuscripts that they believe should exist, but may not actually exist

Seizure and Epilepsy Diagnosis



A pragmatic algorithm to select appropriate antiseizure medications in patients with epilepsy

EpiPick

A Tool for Selecting AntiSeizure Medication

<https://epipick.org>

This antiseizure medication selection tool is designed to assist healthcare providers both in diagnosing / classifying seizures and choosing an appropriate medication (monotherapy). It is designed for use in patients whose seizures start at age 10 years or older.

The algorithm was developed by Ali A. Asadi-Pooya, Sándor Beniczky, Emilio Perucca, Guido Rubboli and Michael R. Sperling. The app was programmed by Stefan Rampp. The project was supported by a grant from the Danish Epilepsy Centre and Filadelfia Research Foundation. The translation into Portuguese was done by Isabella D'Andrea, Vanessa Cristina Colares Lessa, Elza Márcia Yacubian, Katia Lin and Fabio A. Nascimento.



Version 09 June 2022

EpiPick publications

Legal disclaimer

Accept terms

Artificial intelligence (ChatGPT 4.0) vs. Human expertise for epileptic seizure and epilepsy diagnosis and classification in Adults: An exploratory study

Francesco Brigo ^{a,*} , Serena Broggi ^b , Gionata Strigaro ^c, Sasha Olivo ^d,
Valentina Tommasini ^d, Magdalena Massar ^a, Gianni Turcato ^e, Arian Zaboli ^a

- Evaluates ChatGPT's performance in diagnosing and classifying **epileptic seizures, epilepsy, and underlying etiologies** in adult patients compared to epileptologists and neurologists.
- ChatGPT was 'trained' using official ILAE documents on epilepsy diagnosis and classification
- Assessed 37 clinical vignettes based on real adult patient cases
- Reference standard set by a senior epileptologist

Conclusions:

- ✓ ChatGPT demonstrated **high sensitivity** (≥ 96.9 %) in identifying epileptic seizures and diagnosing epilepsy
- ✓ **Lower specificity**, particularly for distinguishing acute symptomatic from unprovoked seizures (33.3 %) and diagnosing epilepsy (26.7 %), leading to **frequent false positives**
- ✓ ChatGPT excelled in diagnosing epileptic syndromes and structural etiologies (accuracy = 90.0 %) but struggled with ambiguous cases such as unknown seizure onset (accuracy = 12.5 %) and rare etiologies



How Intelligent is Artificial Intelligence? Using ChatGPT to Diagnose Epilepsy and Classify Seizures

Epilepsy Currents

1–3

© The Author(s) 2025

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/15357597251345447

journals.sagepub.com/home/epi



Artificial intelligence (ChatGPT 4.0) vs. Human expertise for epileptic seizure and epilepsy diagnosis and classification in Adults: An exploratory study.

Francesco Brigo, Serena Broggi, Gionata Strigaro, Sasha Olivo, Valentina Tommasini, Magdalena Massar, Gianni Turcato, Arian Zaboli. *Epilepsy Behav.* 2025 May;166:110364. doi: 10.1016/j.yebeh.2025.110364. Epub 2025 Mar 12.

Aims: Artificial intelligence (AI) tools like ChatGPT hold promise for enhancing diagnostic accuracy and efficiency in clinical practice. This exploratory study evaluates ChatGPT's performance in diagnosing and classifying epileptic seizures, epilepsy, and underlying etiologies in adult patients compared to epileptologists and neurologists. **Methods:** A prospective simulation study assessed 37 clinical vignettes based on real adult patient cases. ChatGPT was “trained” using official International League Against Epilepsy documents on epilepsy diagnosis and classification. Diagnoses and classifications by ChatGPT, two epileptologists, and two neurologists were compared against a reference standard set by a senior epileptologist. Diagnostic accuracy was evaluated using sensitivity, specificity, positive predictive value, and negative predictive value. Cohen's kappa (κ) was calculated to assess agreement. **Results:** ChatGPT demonstrated high sensitivity ($\geq 96.9\%$) in identifying epileptic seizures and diagnosing epilepsy, ensuring no cases were missed. However, its specificity was lower, particularly for distinguishing acute symptomatic from unprovoked seizures (33.3%) and diagnosing epilepsy (26.7%), leading to frequent false positives. ChatGPT excelled in diagnosing epileptic syndromes ($\kappa = 1.00$) and structural etiologies (accuracy = 90.0%) but struggled with ambiguous cases such as unknown seizure onset (accuracy = 12.5%) and rare etiologies. Human experts consistently outperformed ChatGPT with near-perfect accuracy and higher κ values. **Conclusion:** ChatGPT shows potential as a supplementary diagnostic tool but requires human oversight due to reduced specificity and limitations in nuanced clinical judgment. Further development with diverse datasets and targeted training is necessary to improve AI performance. Integrating AI with expert clinicians can optimize diagnostic workflows in epilepsy care.

Remarks:

- ✓ While their specificity was likewise imperfect (64%) thus potentially overcalling epilepsy despite inclusion of “red flags” historical questions, they found fairly strong agreement between algorithm versus expert seizure classification
- ✓ While the current work had a small sample in a selected number of cases, it still provides an important contribution, likely foreshadowing much research to come as the epilepsy community deliberates exactly what problems we need technology to solve and for whom

Epilepsy Presurgical Decision-Making



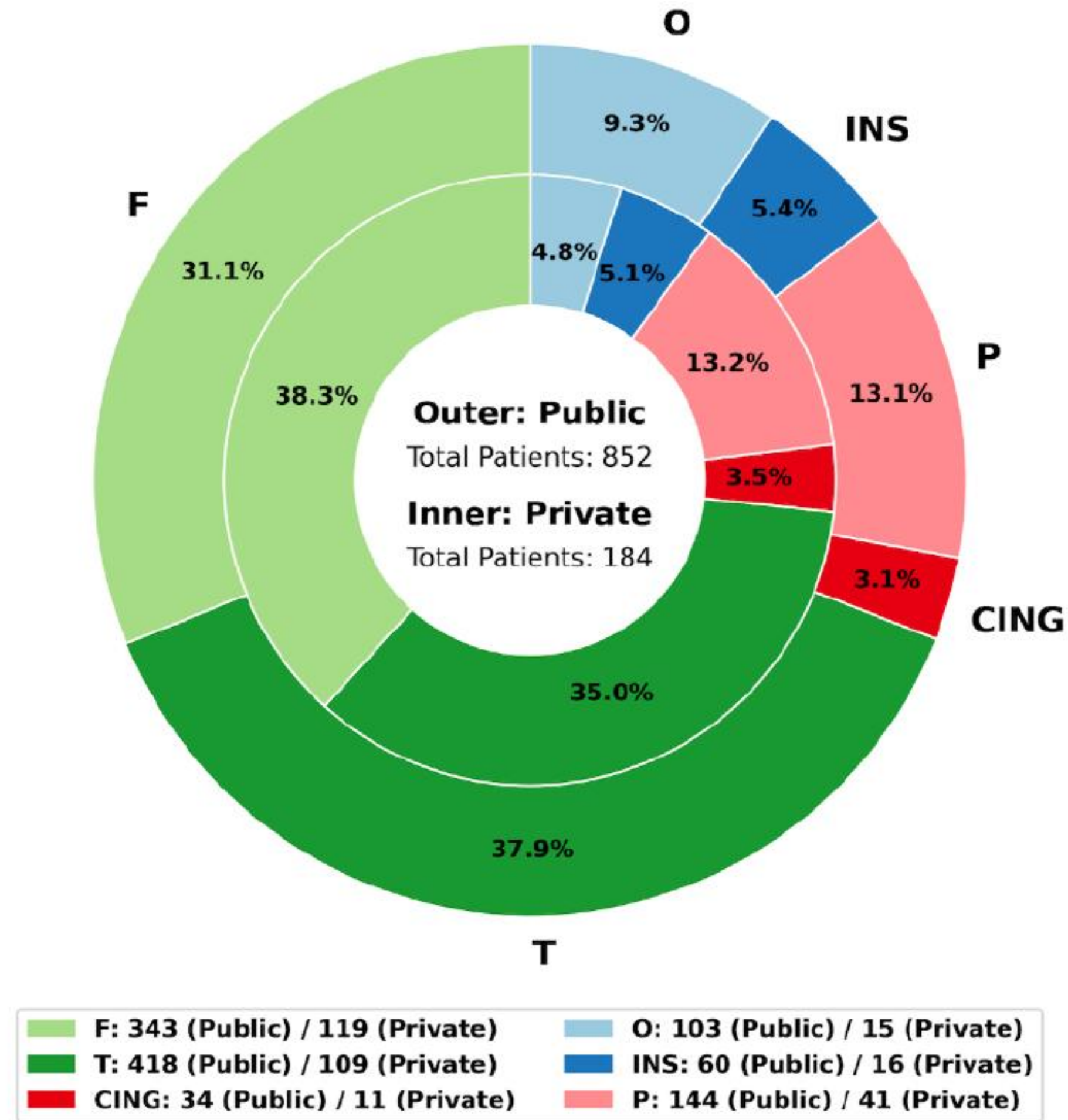
Original Paper

Clinical Value of ChatGPT for Epilepsy Presurgical Decision-Making: Systematic Evaluation of Seizure Semiology Interpretation

Yaxi Luo^{1*}, MS; Meng Jiao^{2*}, PhD; Neel Fotedar^{3,4}, MD; Jun-En Ding², MS; Ioannis Karakis^{5,6}, MD, PhD; Vikram R Rao⁷, MD, PhD; Melissa Asmar⁸, MD; Xiaochen Xian⁹, PhD; Orwa Aboud¹⁰, MD, PhD; Yuxin Wen¹¹, PhD; Jack J Lin⁸, MD; Fang-Ming Hung^{12,13}, MD; Hai Sun¹⁴, MD, PhD; Felix Rosenow¹⁵, MD; Feng Liu^{2,16}, PhD

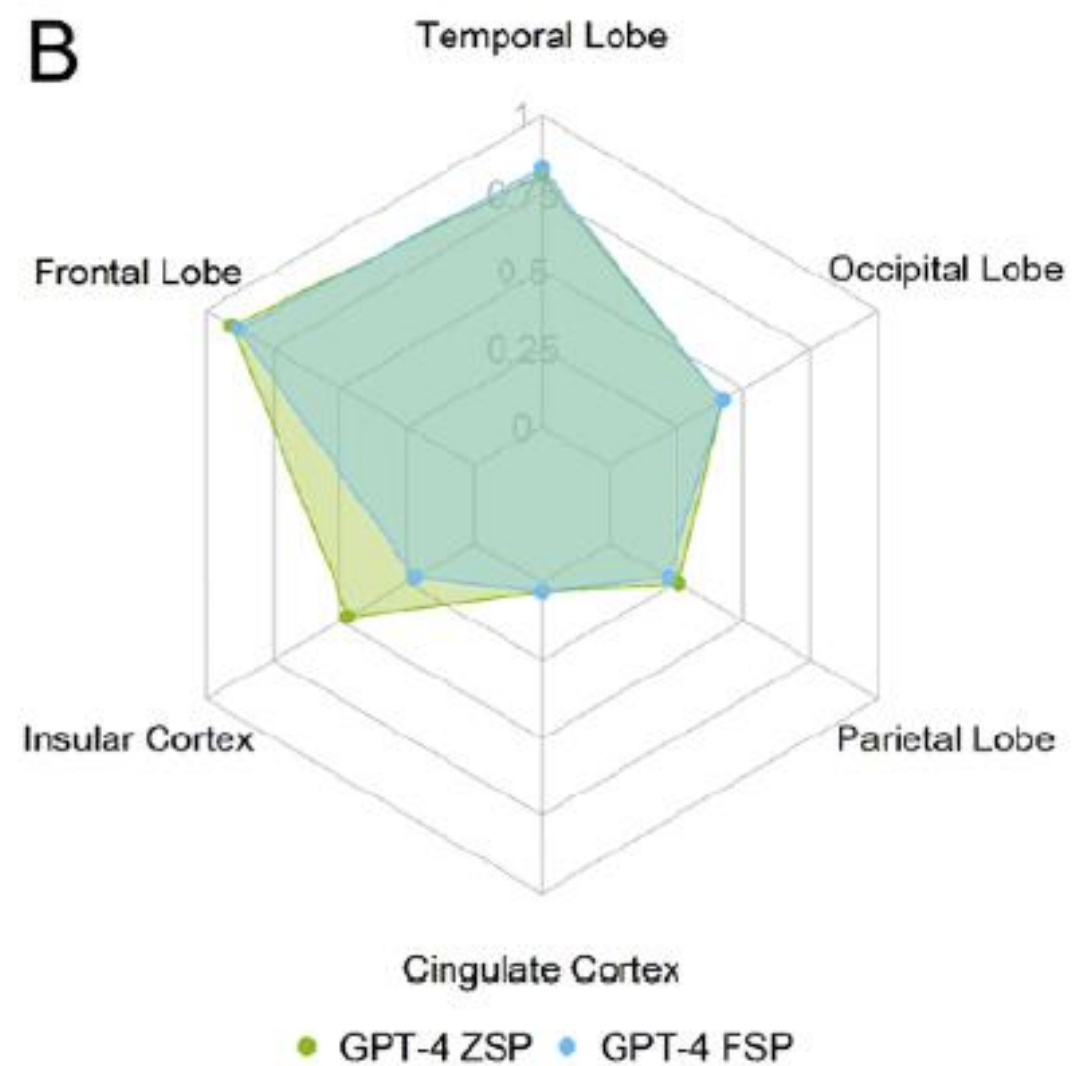
- 2 data cohorts:
 - ❖ **Publicly** sourced cohort of 852 semiology-EZ pairs from 193 peer-reviewed journal publications and
 - ❖ **Private** cohort of 184 semiology-EZ pairs collected from Far Eastern Memorial Hospital (FEMH) in Taiwan

Compare the performance of ChatGPT with 8 Epileptologists

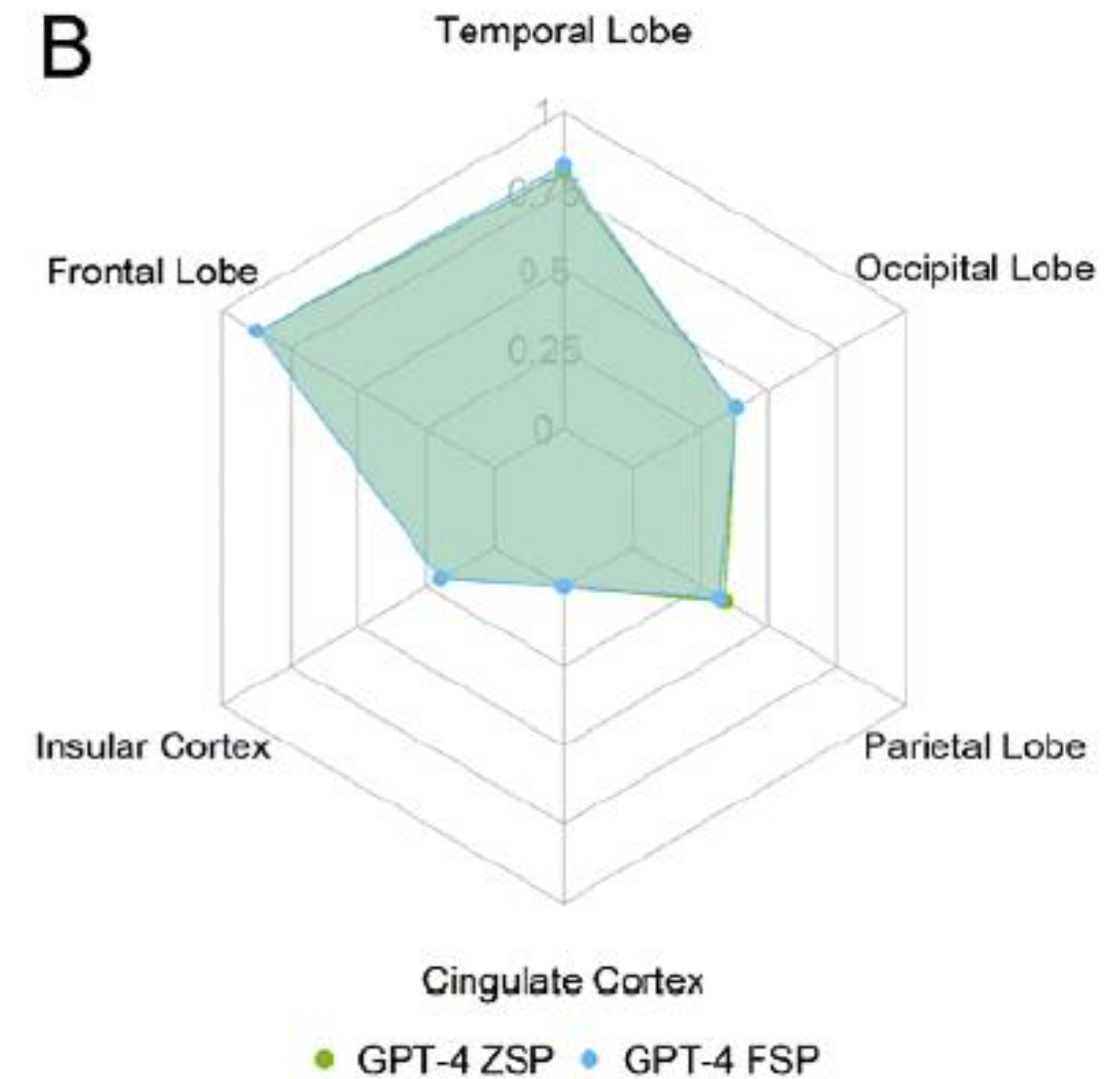


ChatGPT-4

Performance Metric: regional sensitivity (RSens)

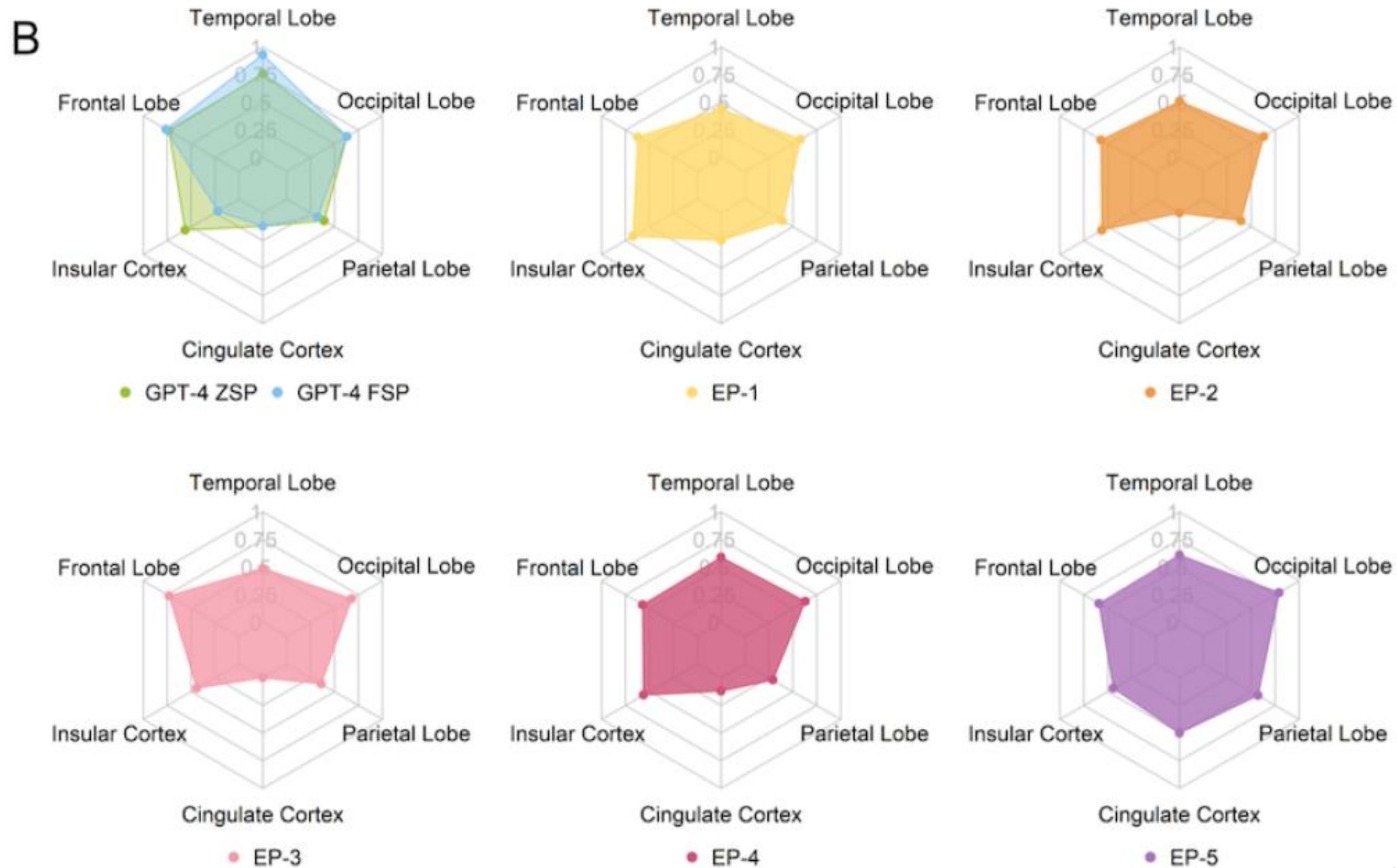


Publicly source data



Privately source data

Performance comparison on a 100-question survey between ChatGPT-4 and five individual epileptologists



Performance comparison on a 100-question survey between ChatGPT-4 and five individual epileptologists

Key messages

- ChatGPT significantly **outperformed** epileptologists at identifying EZ locations in **more common regions**, demonstrated comparable but slightly lower performance in less common regions, and substantially **underperformed** in **rare regions**
- These findings suggest that ChatGPT's performance is positively correlated with the availability of sufficient data to support its responses
- ChatGPT could serve different roles depending on the clinical setting. In **epilepsy centers** with rich resources, it may function as a copilot to support epileptologists at improving diagnostic efficiency. In **resource-limited epilepsy centers**, where access to specialized epilepsy care is scarce, ChatGPT could be particularly valuable for assisting general practitioners or nonspecialist clinicians with preliminary seizure classification and decision-making, potentially improving access to epilepsy care

●●● **Challenges implementing in practice**

- **In practice, EZ localization is not made by a single epileptologist but through consensus in multidisciplinary epilepsy conferences**
- **Note that seizure semiology is just one element contributing to EZ determination**
- **A multimodal approach combining EEG, NPT, sMRI, fMRI, and other data reflects real-world practice and may improve localization accuracy**



Ethical and Practical Challenges

"Black Box" AI Models



AI systems, particularly those based on deep learning, where the decision-making process is opaque and difficult to understand:



"Inequity"



Unfairness or injustice in the distribution of resources, opportunities, or outcomes

Inequity



Disparity

Trust and Security

Societal Issues

"The Digital Divide"

- Unequal access to technology
- Unequal access to the Internet
- Lack of digital literacy
- Lack of support for telehealth visits
- Unequal access to health apps
- Unequal access to technical support



Patients and wearable digital health technology data

Local Health System Strategies

- Health professional–patient engagement
- Patient control over data
- Transparency regarding use of patient data





THANK YOU

FOR YOUR ATTENTION

Chusak.L@Chula.ac.th