2026, Volume 6, Issue 1, Page No: 134-146
ISSN: 3108-4826

Society of Medical Education & Research

Journal of Medical Sciences and Interdisciplinary Research

Predictive Value of Early Amplitude-Integrated and Raw Electroencephalographic Features for Long-Term Neurodevelopment in Extremely Preterm Infants: A Decade-Long Cohort Study from the Netherlands

Lina Hassan[1]*, Omar Khalaf[2], Reem Jaber[1], Rania Mostafa[3]

[1] Department of Neonatology and Neurodevelopment, Faculty of Medicine, University of Jordan, Amman, Jordan.
[2] Department of Preterm Infant Brain Monitoring and EEG, Faculty of Medicine, Jordan University of Science and Technology, Irbid, Jordan.
[3] Department of Long-Term Neurodevelopmental Outcomes, Faculty of Medicine, Al-Balqa Applied University, Salt, Jordan.

***E-mail** ✉ lina.hassan@ju.edu.jo

Abstract

Babies born extremely early, before completing 28 weeks of pregnancy, carry a high chance of facing lasting issues with brain development and function. Monitoring brain activity soon after birth using amplitude-integrated EEG along with the original EEG waveforms (aEEG–EEG) offers promise for forecasting how these infants will develop over time. The goal here was to pinpoint specific qualitative and quantitative aspects of these early recordings that could indicate later neurodevelopmental results. Researchers reviewed data from extremely preterm newborns (delivered before 28 weeks and 0 days) who received standard two-channel aEEG–EEG tracking for the first 72 hours in the neonatal unit at Wilhelmina Children's Hospital, Utrecht, Netherlands, spanning June 2008 through September 2018. Only those free from major genetic conditions, metabolic disorders, or serious congenital disabilities qualified. We extracted a wide range of characteristics from cleaned signals, covering qualitative elements in three main areas—overall background activity, cycles of sleep and wakefulness, and seizure-like events—and quantitative measures across four groups: frequency spectrum details, voltage levels, brain connectivity signals, and periods of signal interruption. Advanced machine learning techniques for both continuous prediction and category sorting tested how well these features could forecast 13 separate developmental measures (involving thinking skills, movement abilities, and behavior concerns) when children reached 2–3 years and again at 5–7 years. All evaluations adjusted for influencing variables including exact weeks of gestation, mother's schooling, how sick the baby was, total morphine given, serious brain damage, and any drugs for seizures, calming, or anesthesia. The analysis covered 369 eligible infants and generated 339 distinct aEEG–EEG variables (9 qualitative plus 330 quantitative). Predictive regression models built with machine learning reached meaningful but limited success (correlations between 0.13 and 0.23) across nine of the 13 endpoints. Classification models, however, performed reasonably well at spotting children who later showed intellectual challenges versus those with the strongest results at school-entry age. Balanced accuracy hit 0.77 [95% CI: 0.62–0.90; P = 0.0020] for overall intelligence scores and 0.81 [0.65–0.96; P = 0.0010] for language-based intelligence scores. These same accuracy levels held steady even after limiting inputs strictly to the quantitative variables. Results show that brain monitoring data collected right after birth can help automatically flag extremely preterm babies at risk for difficult developmental trajectories. Such tools could become useful, understandable aids for clinicians making decisions and shaping care plans.

Keywords: Extremely preterm infants, Amplitude-integrated EEG, Raw EEG, Neurodevelopmental outcomes, Machine learning, Neonatal brain monitoring

Access this article online https://smerpub.com/

Received: 22 December 2025; **Accepted:** 29 March 2026

How to cite this article: Hassan L, Khalaf O, Jaber R, Mostafa R. Predictive Value of Early Amplitude-Integrated and Raw Electroencephalographic Features for Long-Term Neurodevelopment in Extremely Preterm Infants: A Decade-Long Cohort Study from the Netherlands. J Med Sci Interdiscip Res. 2026;6(1):134-46. https://doi.org/10.51847/q5Lwv58DNK

Introduction

Thanks to major progress in pregnancy and newborn medical support during recent decades, far more infants born well before 28 weeks now survive [1-3]. Even so,

many still encounter significant ongoing problems with cognitive abilities and physical coordination [4, 5].

These difficulties usually trace back to harm or irregularities affecting the immature brain while the baby grows outside the protective womb environment [6]. Providing focused protection and quick support during the NICU stay is therefore crucial for minimizing future disabilities [7]. This drives strong interest in identifying reliable early signals from brain function that can predict longer-term outcomes while the infant remains under hospital care [8].

aEEG paired with raw EEG traces has grown popular globally as a convenient, real-time method for watching brain performance in intensive care settings for newborns [9, 10]. Offering a simpler setup than full EEG systems, it needs just one or two recording channels and is quick to apply [10]. Monitoring can start at admission, supporting a fast response to emerging brain issues. Both descriptive and measured parameters from these recordings have repeatedly shown value for estimating long-range development in preterm babies [9, 11-18]. When compared with MRI scans, aEEG–EEG delivers similar forecasting power [19] and also identifies at-risk infants who lack obvious structural damage on imaging [20].

Despite expanding studies on this topic, important questions persist, especially around the most immature babies [13]. Earlier projects often folded them into larger preterm samples with fairly small numbers overall. Because brain-signal patterns shift noticeably week to week in the preterm phase [21], observations from older or less premature groups do not transfer cleanly.

Inconsistent study designs and evaluation methods further complicate comparisons [22]. Outcome checks happen at different ages and do not always employ the same validated tools. Recording windows also differ. Focusing on signals from the earliest days could uncover the most timely warning signs.

Past efforts typically focused on single feature types instead of testing how different categories might work together [23]. Modern machine learning now makes it practical to handle many features at once. This moves beyond older group-based statistical links that were standard before, toward personalized forecasts needed for future tailored medical approaches [24].

Accordingly, this investigation tested whether brain monitoring information from the first three days could signal developmental prospects at toddler and early-school stages within a sizable, uniform set of extremely preterm infants. Comprehensive qualitative plus quantitative features fed into machine learning models for both regression-style association testing and classification to separate good versus impaired outcomes, aiming for direct usefulness in medical practice.

Materials and Methods

Study design and population

This backward-looking group analysis focused on extremely premature newborns (delivered before 28 weeks and zero days) who entered the neonatal intensive care unit at Wilhelmina Children's Hospital (WKZ) in Utrecht, Netherlands, from June 1, 2008, until September 30, 2018. Inclusion was limited to those who underwent continuous two-channel aEEG–EEG tracking in the first three days after birth and who did not have genetic disorders, metabolic conditions, or significant congenital abnormalities. The Medical Research Ethics Committee at the University Medical Center Utrecht approved data use [protocol number 20-660-C]. Because the information originated from routine clinical practice and the review was retrospective, no signed parental consent was required. All records were pseudonymized and stripped of identifying details before processing.

Procedures

In line with routine protocols for the most premature admissions at WKZ's NICU, two-channel aEEG–EEG observation began promptly upon arrival and continued at the bedside throughout the first three days of life. Recordings were performed with the BrainZ monitor (models BRM2 or BRM3; Natus Medical, Seattle, WA, USA). Original EEG waveforms came from needle electrodes positioned under the skin above the frontoparietal brain regions (left side: F3-P3; right side: F4-P4), captured at a 256 Hz sampling frequency, and then converted into aEEG displays. A separate reference electrode sat at the top of the head (Cz position).

Trained specialists (CT, LW, and MLT) initially examined all aEEG–EEG recordings via Analyze Research software (version 2.0, BrainZ Instruments). For each of the three days, experts hand-picked the optimal one-hour segment showing the most developed activity and minimal interference. These windows were set at 20–24 hours for the first day, 44–48 hours for the second, and 68–72 hours for the third. This selected high-quality hour provided data for both descriptive background evaluations and numerical feature computations, as those

measurements are particularly vulnerable to noise and disturbances (**Figure 1**).

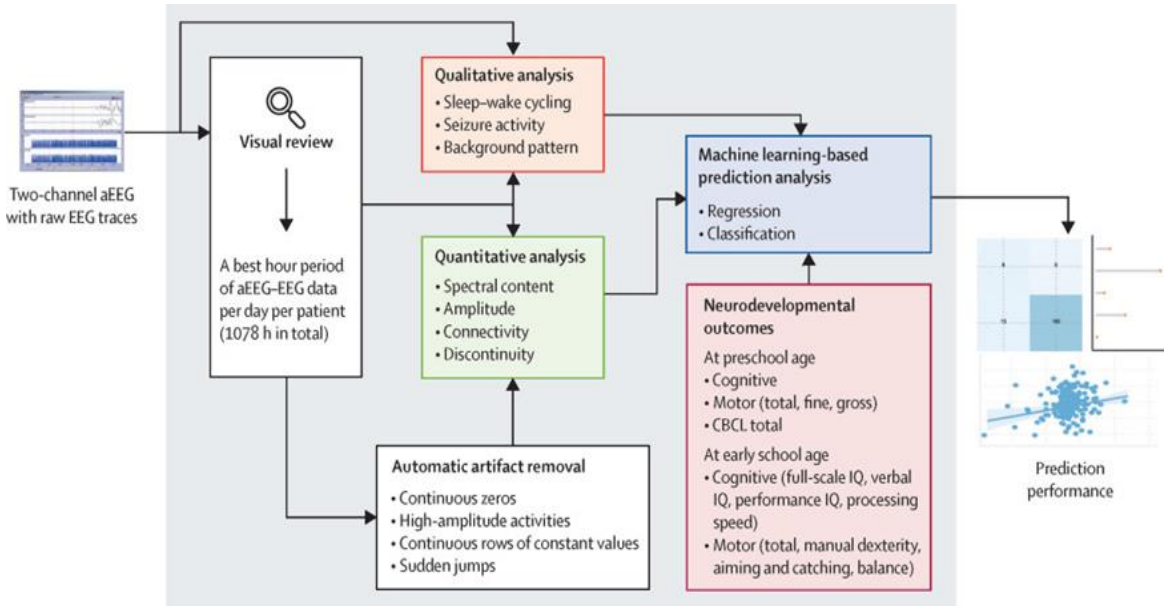


Figure 1. Schematic representation of the full processing workflow. Abbreviations: aEEG = amplitude-integrated electroencephalogram. CBCL = Child Behavior Checklist. EEG = electroencephalogram. IQ = intelligence quotient.

Clinicians with expertise (CT, LW, and MLT) performed additional descriptive evaluations of the aEEG–EEG recordings according to the Hellström-Westas framework [25]. To promote consistency and dependability, at least two reviewers assessed the traces together for each evaluation. They remained unaware of the infants' clinical details and later developmental outcomes throughout the review.

Descriptive assessments covered three main categories: overall background activity, patterns of sleep and alertness cycles, and the presence of seizure-like events. For background activity review, the chosen optimal one-hour segment from each day was divided into six equal 10-minute intervals. The dominant category across those intervals was retained for downstream use. Reviews of sleep–wake cycles and seizure activity drew on the full daily recordings.

Many of these characteristics included several subcategories. Because certain subtypes appear infrequently among extremely preterm babies, the team simplified the results into a straightforward two-level format—typical or atypical—for clarity and to emphasize key signals.

Numerical analysis used the original EEG waveforms and custom MATLAB scripts (MathWorks, Natick, MA, USA) developed internally, supplemented by the NEURAL MATLAB toolbox [26].

Before feature extraction, the selected raw EEG segments underwent automated cleaning to remove artifacts and reduce noise. Portions showing clear disturbances—such as flat zero lines, extremely large voltage swings, unchanging value sequences, or abrupt shifts—were discarded. A bandpass filter (0.5–30 Hz) then removed unwanted low- and high-frequency components, followed by a 50 Hz notch filter to suppress electrical line noise.

This cleaned data yielded 110 numerical characteristics, organized into four groups: spectral content (26 items), amplitude (56 items), connectivity (21 items), and discontinuity (7 items).

Both descriptive and numerical aEEG–EEG characteristics were derived separately for each of the three days.

Outcomes

Standard follow-up appointments at the WKZ outpatient department provided neurodevelopmental assessments

for the extremely preterm cohort at preschool age (2–3 years) and early school age (5–7 years). Evaluators remained blinded to the early aEEG–EEG findings.

At younger ages, the Bayley Scales of Infant and Toddler Development, Third Edition (BSID-III) [27] measured cognitive and motor performance, providing four key indices: cognitive composite, overall motor composite, and separate fine and gross motor subscale scores. The Child Behavior Checklist (CBCL) gauged behavioral and emotional issues, producing a combined problem score [28].

At early school age, the Wechsler Preschool & Primary Scale of Intelligence, Third Edition (WPPSI-III) [29] evaluated thinking skills through full-scale IQ plus verbal IQ, performance IQ, and processing speed composites. Motor skills were assessed using the Movement Assessment Battery for Children, Second Edition (MABC-2) [30], which produced a total motor score and three component scores for manual dexterity, aiming/catching, and balance.

All tests were administered in Dutch. These are all standardized instruments with population-based norms. Lower results on BSID-III, WPPSI-III, and MABC-2 reflect reduced abilities, while elevated CBCL scores signal greater behavioral or emotional difficulties. For grouping purposes, every outcome was turned into a two-category variable (strong versus compromised) using cutoffs at ±2 standard deviations from the expected average. On scales with a mean of 100 (SD: 15), such as BSID-III composites and WPPSI-III scores, values of 70 or below were counted as compromised. In particular, a WPPSI-III IQ of 70 or lower at 5–7 years denotes intellectual disability. On scales with a mean of 10 (SD: 3), such as BSID-III subscales and MABC-2 scores, results of 4 or below were considered compromised. For the CBCL total problem score (mean 50, SD 10), a value of 70 or above indicated impairment.

Statistical analysis

Multiple infant-specific variables can influence links between aEEG–EEG signals and later development. To isolate the genuine forecasting strength of the brain recordings, we incorporated key adjusting factors: gestational age at delivery, mother's educational background, overall sickness level, total morphine received, occurrence of major brain injury, and exposure to seizure-control, calming, or anesthesia medications. A Spearman rank correlation matrix (rS) illustrated associations among aEEG–EEG characteristics, these adjusting factors, and developmental measures. Before modeling, each aEEG–EEG characteristic was corrected for the adjusting factors via ordinary least-squares regression. Missing data points were filled using median values for numeric items and mode values for categorical ones.

For every developmental measure, the team constructed a support vector regression model and a histogram-based gradient boosting classification model. We trained models with nested 3-fold cross-validation, using grid search to identify the best settings. Classification folds were stratified to preserve class balance. Classification models were developed only for outcomes having at least nine cases in each category, guaranteeing sufficient examples per inner fold for reliable tuning.

Final performance was tested using leave-one-subject-out cross-validation with the optimized settings. Regression performance was quantified by Pearson correlation (r) and mean squared error (MSE) between observed and forecasted scores. The r value spans –1 to 1 (1 = perfect agreement, 0 = random-level, negative = worse than chance). MSE ranges from 0 (ideal) upward (higher = poorer fit). Classification performance relied on balanced accuracy and Fβ score, both ranging from 0 (incorrect) to 1 (correct). Balanced accuracy of 0.5 matches random guessing. The Fβ score, with β set to 10, emphasizes sensitivity (recall) more than precision through a weighted harmonic average.

Permutation testing (1,000 random shuffles of outcome labels) assessed whether results exceeded chance levels. The P-value was computed as $(n + 1)/1001$, where n equals the number of shuffled runs that outperformed the true data. The threshold for significance was $P < 0.05$. For regression, both r and MSE needed to be significant; for classification, both balanced accuracy and Fβ score needed to be significant. We derived 95% confidence intervals from 1,000 bootstrap samples. Shapley additive explanations (SHAP) values further quantified the relative importance of feature groups and daily contributions, with larger SHAP figures showing stronger influence.

Summary statistics and graphics were generated in Python (version 3.10.9) and R (version 4.2.3) through RStudio (version 2023.03.0+386). Machine learning steps used scikit-learn (version 1.2.1) in Python (version 3.10.9).

Results and Discussion

Screening identified 601 extremely preterm infants. Exclusion occurred for 108 due to missing aEEG–EEG recordings or presence of major congenital anomalies, and for 124 because of inadequate, invalid, or low-quality aEEG–EEG data. This left 369 infants who satisfied all criteria and entered the final analysis. **Table 1** presents their background details and clinical course in the NICU.

Table 1. Baseline characteristics of the study population.

Category	Variables	Extremely preterm infants (n = 369)
Maternal and infant demographic data	**Sex at birth***	
	Female	164 (44%)
	Male	205 (56%)
	Gestational age at delivery, weeks	26.4 (1.1; 23.9–27.9)
	Birth weight†, g	880 (180)
	Maternal educational attainment	
	No formal education	1 (< 1%)
	Primary schooling	11 (3%)
	Incomplete secondary education	37 (10%)
	Completed secondary education	80 (22%)
	Higher education/university degree	96 (26%)
	Not reported	144 (39%)
NICU-related clinical features	**Morphine exposure during hospitalization**	
	Received morphine	222 (60%)
	Did not receive morphine	142 (39%)
	Missing information	5 (1%)
	Total morphine exposure‡, mg/kg	1.3 (2.3)
	Exposure to antiseizure, sedative, or anesthetic agents§	
	Yes	140 (38%)
	No	229 (62%)
	Severity of illness	
	Severe	183 (50%)
	Mild	181 (49%)
	Missing data	5 (1%)
	Severe brain injury status	
	Present	118 (32%)
	Absent	246 (67%)
	Missing data	5 (1%)
	Apgar assessment	
	1 minute after birth¶	5 (3–7)
	5 minutes after birth‖	8 (6–8)
	10 minutes after birth**	8 (8–9)
Neurodevelopmental follow-up assessments	Age at BSID-III evaluation, years	2.5 (0.2; 2.2–3.0)
	Age at CBCL evaluation, years	2.5 (0.2; 2.2–3.0)
	Age at WPPSI-III evaluation, years	5.9 (0.2; 5.1–6.8)
	Age at MABC-2 evaluation, years	5.9 (0.2; 5.1–7.0)

Data presentation notes: Values are shown as n (%), mean (SD; range), mean (SD), or median (IQR). BSID-III = Bayley Scales of Infant and Toddler Development, Third Edition. CBCL = Child Behavior Checklist. MABC-2 = Movement Assessment Battery for Children, Second Edition. NICU = Neonatal Intensive Care Unit. WPPSI-III = Wechsler Preschool & Primary Scale of Intelligence, Third Edition.

*The terms “female” and “male” indicate sex assigned at birth.

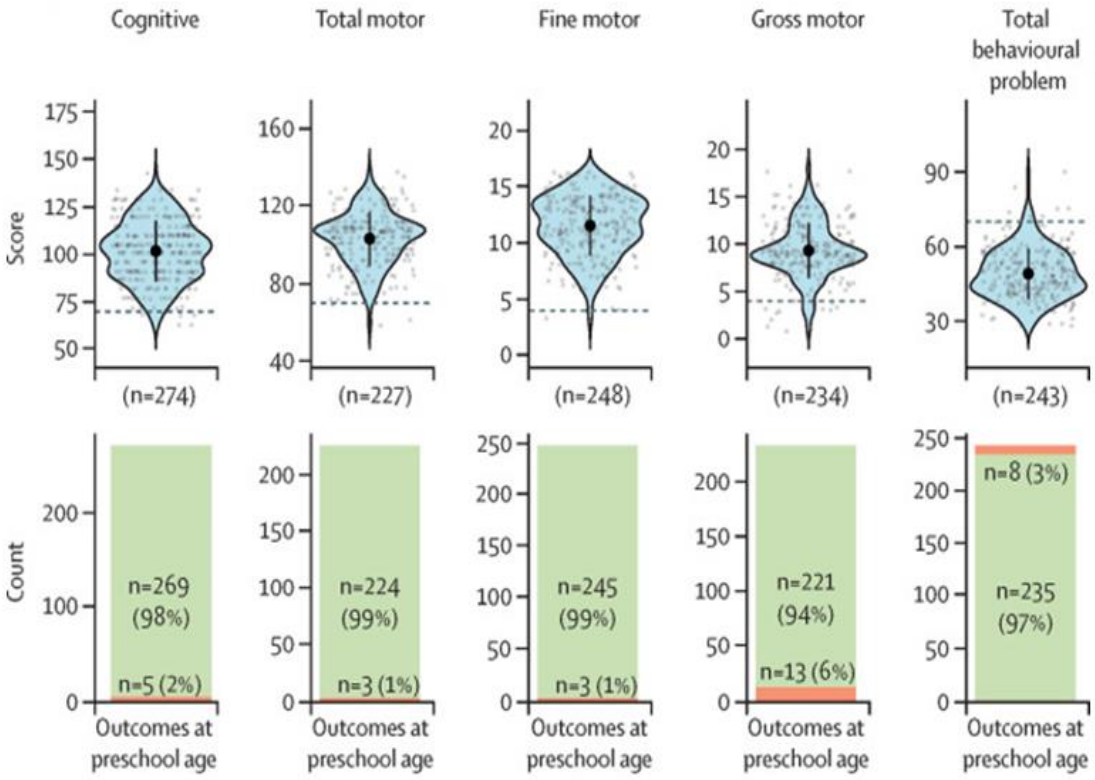
†Data unavailable for one infant.

‡Data unavailable for 11 infants.

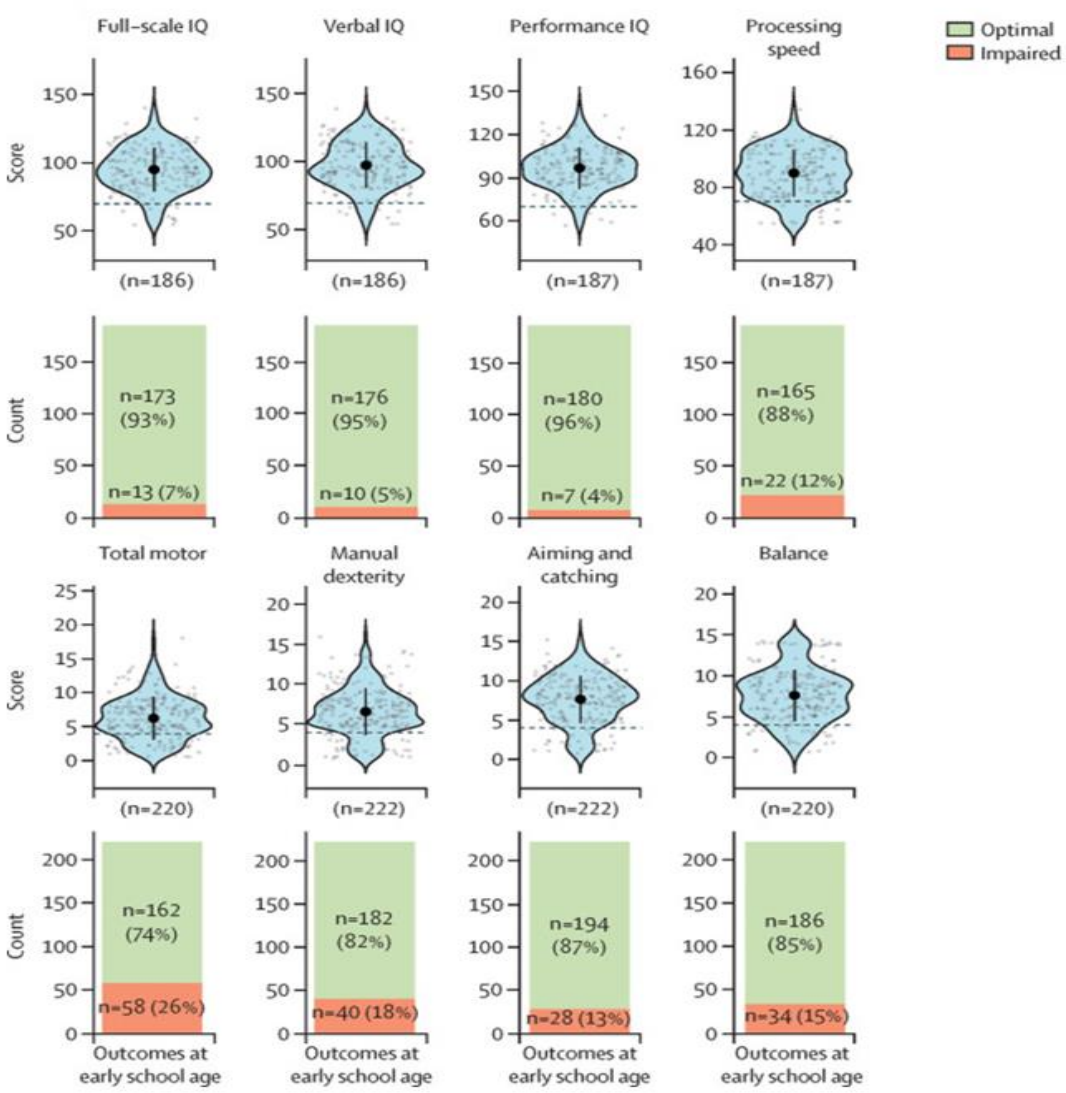
§Antiseizure, calming, or anesthesia medications listed here encompass phenobarbital, lidocaine, levetiracetam, clonazepam, midazolam, along with any additional agents possibly used for surgical anesthesia.
¶Data unavailable for seven infants.
‖Data unavailable for four infants.
**Data unavailable for 86 infants.

Among the 369 participants, 343 infants (93%) possessed usable aEEG–EEG recordings across all three targeted windows (20–24 h, 44–48 h, and 68–72 h). The rest had information for only one period (three infants, 1%) or two periods (23 infants, 6%). Overall, 1078 hours of reasonably high-quality aEEG–EEG recordings were collected and processed for characteristic extraction (**Figure 1**). For each infant, researchers derived a full set of 339 aEEG–EEG characteristics, consisting of nine descriptive (qualitative) and 330 numerical (quantitative) characteristics.

Of the 369 infants, outcome data at preschool age were available for 227 (62%; total motor score) to 274 (74%; cognitive score) children. At early school age, available data ranged from 186 (50%; full-scale IQ) to 222 (60%; aiming and catching) children (**Figure 2**). Impaired performance at preschool age affected 3 (1%) to 13 (6%) infants depending on the specific measure. At early school age, the proportion showing impairment varied from 7 (4%) to 58 (26%). **Table 1** provides the exact ages at which these assessments took place.



(a)



(b)

Figure 2. Spread of neurodevelopmental scores at preschool age (a) and early school age (b).

In the violin plots, the black dot in the middle marks the average value, the vertical line through it shows the standard deviation, and the horizontal dashed line marks the cutoff set at ±2 standard deviations from the population norm. The count of infants with available measurements for each outcome appears beneath the corresponding plot. IQ = intelligence quotient.

A 358 × 358 matrix of Spearman rank correlation coefficients is displayed in **Figure 3**. It depicts the pairwise associations among the 339 aEEG–EEG characteristics, the confounding variables, and the 13 developmental outcomes. Links among the aEEG–EEG characteristics themselves tended to be stronger than connections across different categories. Associations between single aEEG–EEG characteristics and the outcomes proved generally modest (absolute values below 0.29). The confounding factors also showed clear links to both the aEEG–EEG characteristics and the outcomes, highlighting their possible influence on the connections between early brain recordings and later development.

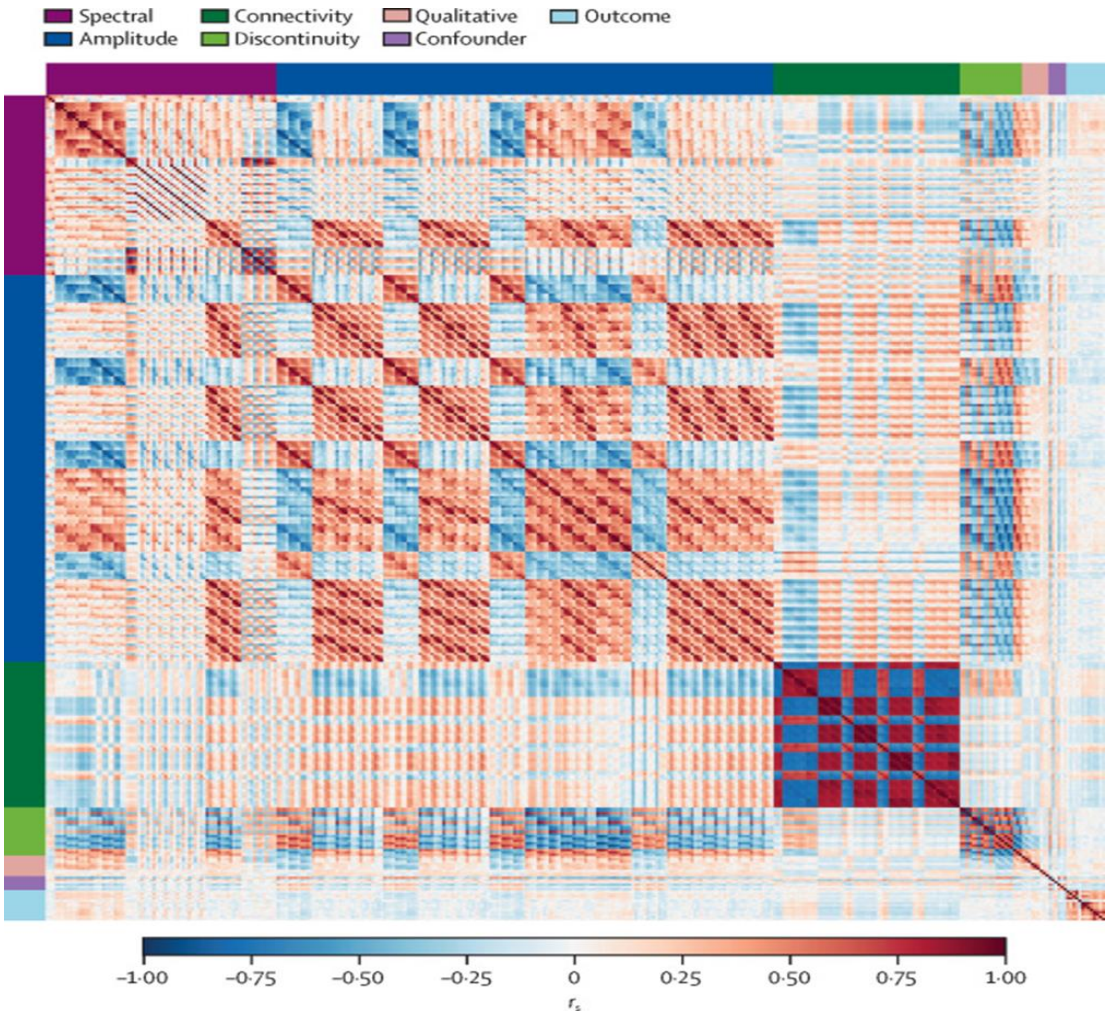


Figure 3. Spearman rank correlation matrix displaying relationships among aEEG–EEG characteristics, adjusting factors, and developmental outcomes.

Various categories of aEEG–EEG characteristics and outcome measures are highlighted using distinct colors, as indicated by the color keys along the top and left edges of the matrix. Inside the matrix, red indicates positive Spearman correlation coefficients (rS), while blue indicates negative values. Color depth reflects correlation strength—darker tones signify stronger associations, and paler tones indicate weaker ones. Abbreviations: aEEG = amplitude-integrated electroencephalogram, and EEG = electroencephalogram.

Machine learning regression models showed that aEEG–EEG characteristics could significantly predict nine of the 13 developmental measures (**Table 2**). Even so, these models remained only moderately predictive, with correlation values ranging from r = 0.13 to r = 0.23.

Table 2. Performance of machine learning regression approaches in predicting neurodevelopmental outcomes.

Assessment time point	Neurodevelopmental measure	Predicted outcome variable	Correlation coefficient (r)	Permutation p-value for correlation (r)	Mean squared error (MSE)	Permutation P-value for MSE
Preschool-age outcomes	**Bayley Scales of Infant and Toddler Development, Third Edition (BSID-III)**	Cognitive composite score	0.14	0.013	38.72	0.013
		Overall motor composite score	0.10	0.056	57.32	0.058

		Fine motor scaled score	0.13	0.014	19.95	0.010
		Gross motor scaled score	0.08	0.13	1.08	0.055
	Child Behavior Checklist (CBCL)	Total behavioral difficulties score	0.13	0.023	7.55	0.022
Early school-age outcomes	**Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III)**	Full-scale intelligence quotient (IQ) score	0.22	0.0050	1.09	0.013
		Verbal IQ score	0.23	0.0030	1.14	0.040
		Performance IQ score	0.19	0.0080	1.17	0.057
		Processing speed score	0.18	0.017	1.11	0.019
	Movement Assessment Battery for Children, Second Edition (MABC-2)	Total motor ability score	0.20	0.0010	4.75	0.0010
		Manual dexterity subscore	0.21	0.0020	2.27	0.0030
		Ball skills (aiming and catching) subscore	0.11	0.055	2.81	0.064
		Balance subscore	0.17	0.0060	6.40	0.0060

Eight developmental measures contained enough cases in each group (favorable versus affected) to allow dependable tuning of model parameters for the classification phase (**Figure 2**). Within this set, regression modeling using aEEG–EEG characteristics did not yield significant results for the gross motor or aiming-and-catching scores (**Table 2**). Consequently, we removed these two measures from additional classification efforts, yielding six classification models (**Table 3**). Models focused on full-scale IQ and verbal IQ showed statistically meaningful results (**Table 3**; **Figure 4a**).

Table 3. Prediction performance of machine learning-based classification models

	Balanced accuracy (95% CI)	**Permutation P-value for balanced accuracy**	**Fβ score (95% CI)**	**Permutation P-value for Fβ score**	**Precision (95% CI)**	**Recall (95% CI)**
			WPPSI-III			
Full-scale IQ score	0.77 (0.62–0.90)	0.0020	0.61 (0.33–0.88)	0.0020	0.38 (0.17–0.61)	0.62 (0.33–0.89)
Verbal IQ score	0.81 (0.65–0.96)	0.0010	0.69 (0.37–0.98)	0.0010	0.35 (0.14–0.56)	0.70 (0.38–1.00)
Processing speed score	0.55 (0.47–0.63)	0.095	0.18 (0.04–0.35)	0.089	0.22 (0.05–0.43)	0.18 (0.04–0.35)
			MABC-2			
Total motor score	0.51 (0.46–0.57)	0.31	0.21 (0.10–0.32)	0.31	0.29 (0.16–0.44)	0.21 (0.10–0.32)
Manual dexterity score	0.54 (0.48–0.61)	0.11	0.20 (0.08–0.33)	0.12	0.28 (0.11–0.44)	0.20 (0.08–0.33)
Balance score	0.52 (0.46–0.59)	0.27	0.15 (0.04–0.27)	0.24	0.20 (0.05–0.37)	0.15 (0.04–0.27)

The Fβ score was calculated with β = 10. Abbreviations: IQ = intelligence quotient, MABC-2 = Movement Assessment Battery for Children, Second Edition. WPPSI-III=Wechsler Preschool and Primary Scale of Intelligence, Third Edition.

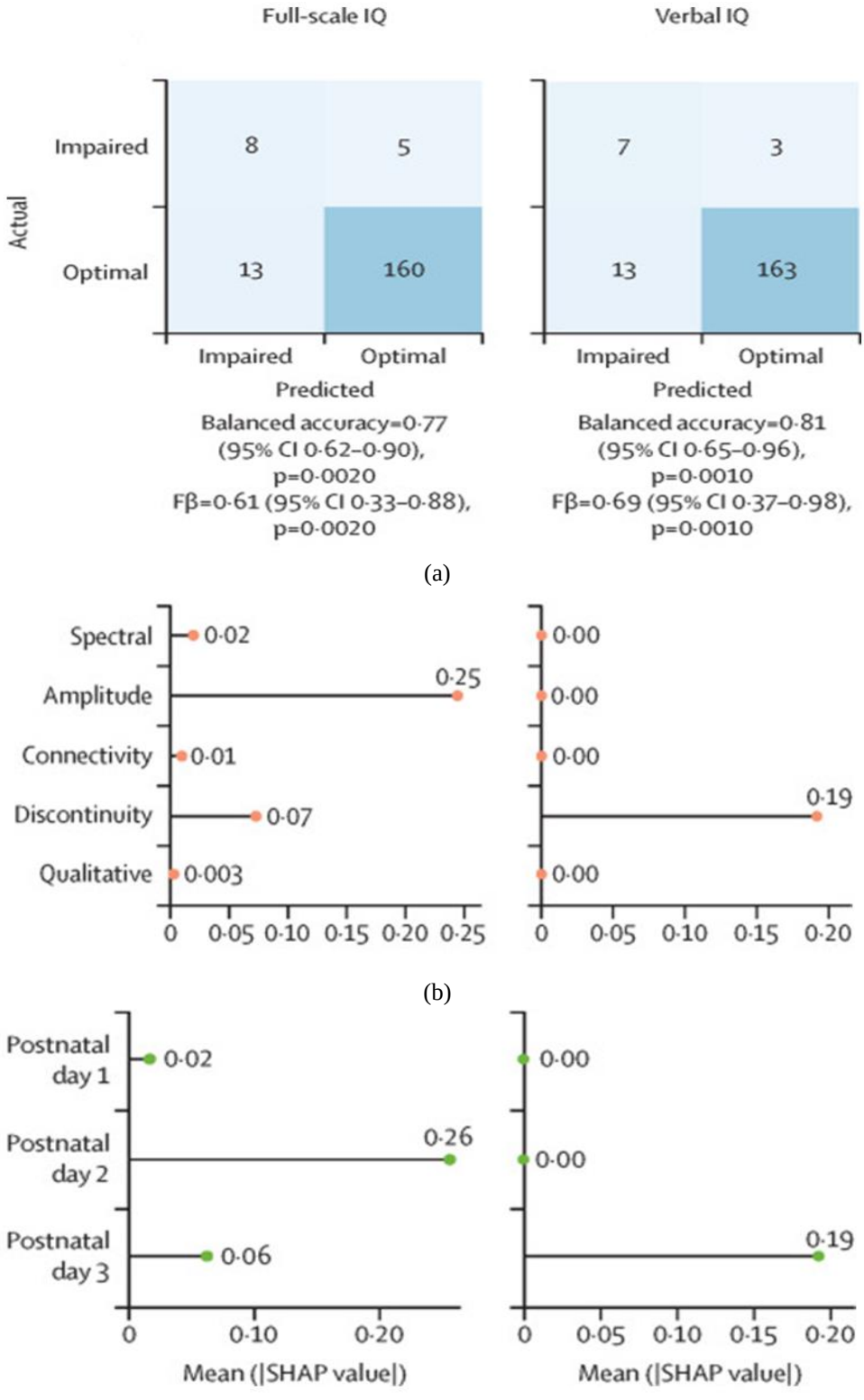


Figure 4. Performance of classification models for full-scale IQ and verbal IQ

Confusion matrices showing predicted versus observed categories (strong versus affected; panel A). These matrices present the numbers of accurate and mistaken classifications, with shading intensity indicating frequency—lighter tones for fewer instances and darker tones for more frequent ones. Feature group contributions (panel B) and daily contributions (panel C) are illustrated through SHAP values. Larger absolute SHAP values reflect greater influence on the classification decisions. IQ = intelligence quotient. SHAP = Shapley additive explanations.

Amplitude-related characteristics contributed most to the full-scale IQ classification model by average absolute SHAP values, followed by discontinuity, spectral content, connectivity, and descriptive features (**Figure 4b)**. Features recorded on day 2 had the largest overall impact for full-scale IQ, followed by those from day 3 and day 1 (**Figure 4c**). For the verbal IQ classifier, only discontinuity characteristics from day 3 played a notable role in correct predictions.

Additional classifiers were developed using exclusively numerical (quantitative) features to forecast full-scale and verbal IQ. These performed at the same level as models that included both descriptive and numerical features: for full-scale IQ, balanced accuracy reached 0.77 (95% CI: 0.63–0.91, P = 0.0030) and Fβ score was 0.61 (95% CI: 0.33–0.88, P = 0.0030); for verbal IQ, balanced accuracy was 0.81 (95% CI: 0.65–0.96, P = 0.0010) and Fβ score was 0.69 (95% CI: 0.37–0.98, P = 0.0010).

This investigation assessed how well descriptive and numerical characteristics from two-channel aEEG–EEG recordings in the first three days after birth could forecast extended neurodevelopmental trajectories in a sizable group of extremely preterm newborns. Findings indicated that these characteristics offered statistically notable yet modest predictive strength for nine of the 13 outcomes in machine learning regression models. However, machine-learning classification approaches effectively separated infants who later showed intellectual disabilities from those who did not, with the strongest results at early school age, achieving balanced accuracies of 0.77 for full-scale IQ and 0.81 for verbal IQ.

These outcomes highlight the practical value of aEEG–EEG data gathered shortly after birth. Unlike earlier investigations that involved limited participant numbers [14-17], this work included 369 extremely preterm infants, strengthening the reliability of the conclusions. The broad extraction of various characteristics provided the first detailed overview of aEEG–EEG patterns in this population and clarified their interconnections. The extensive feature collection further enabled the creation of machine learning systems capable of individualized forecasts, supporting advances toward precision medical care [8, 23].

The observation that classifiers relying solely on numerical features matched the performance of those combining both descriptive and numerical features holds important practical significance. Numerical aEEG–EEG characteristics offer advantages in being objective, straightforward to obtain, and quickly available compared with descriptive ones [13, 23]. This supports the potential to build a fully automated bedside system that delivers prognostic insights, helping medical teams choose care intensity and allocate resources. For instance, such predictions could enable early recognition of infants at elevated risk, allowing prompt protective strategies for the developing brain and better long-term outcomes.

SHAP values helped rank how much different characteristics influenced the classification models. A SHAP value reflects a feature's role within a specific model rather than its absolute real-world importance. Because model predictions are not always accurate, SHAP interpretations may not perfectly mirror true relationships. Thus, even though descriptive features showed smaller contributions to the IQ models here, their possible usefulness for refining numerical measures or in other clinical contexts should not be overlooked.

It should be emphasized that an extremely preterm infant's developmental path is not determined only by the first few days. Repeated predictions over time could better support positive development. Longitudinal modeling while the infant remains in the NICU would permit ongoing adjustments to care strategies. As the baby grows, additional information from clinical observations, other brain monitoring methods, and imaging approaches (including MRI, cranial ultrasound, and near-infrared spectroscopy) becomes accessible. Combining aEEG–EEG data with these sources could yield more dependable forecasts. Additional studies are required to test such integrated approaches.

Limitations of this research include its single-center design. Despite the relatively large sample (n = 369) and use of cross-validation to bolster reliability, external validation with separate cohorts would be valuable. Another issue is the difficulty of accounting for every

possible influencing factor on the EEG–outcome link, particularly elements arising after the NICU period. Furthermore, while two-channel aEEG–EEG is convenient, it captures only limited aspects of brain activity, which may constrain predictive accuracy. Standard multichannel EEG supplies richer data but presents practical difficulties in extremely preterm infants due to small head size and delicate skin. Future efforts might explore new methods, such as dry electrodes, to enable multichannel recordings in the earliest days.

Conclusion

In conclusion, two-channel aEEG–EEG characteristics obtained during the first three postnatal days demonstrated meaningful value for anticipating long-term neurodevelopmental results in a large cohort of extremely preterm infants. Machine learning models enabled, for the first time, the integration and direct comparison of an extensive range of descriptive and numerical aEEG–EEG characteristics. Numerical features displayed stronger predictive capability than descriptive ones. These results support the development of an automated prognostic instrument for identifying long-term risks, which could inform personalized early interventions in the neonatal intensive care unit.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Costeloe KL, Hennessy EM, Haider S, Stacey F, Marlow N, Draper ES. Short term outcomes after extreme preterm birth in England: comparison of two birth cohorts in 1995 and 2006 (the EPICure studies). BMJ. 2012;345:e7976.
2. Glass HC, Costarino AT, Stayer SA, Brett CM, Cladis F, Davis PJ. Outcomes for extremely premature infants. Anesth Analg. 2015;120(6):1337-51.
3. Stoll BJ, Hansen NI, Bell EF, Walsh MC, Carlo WA, Shankaran S, et al. Trends in care practices, morbidity, and mortality of extremely preterm neonates, 1993–2012. JAMA. 2015;314(10):1039-51.
4. Rogers EE, Hintz SR. Early neurodevelopmental outcomes of extremely preterm infants. Semin Perinatol. 2016;40(8):497-509.
5. Twilhaar ES, Wade RM, de Kieviet JF, van Goudoever JB, van Elburg RM, Oosterlaan J. Cognitive outcomes of children born extremely or very preterm since the 1990s and associated risk factors: a meta-analysis and meta-regression. JAMA Pediatr. 2018;172(4):361-7.
6. Volpe JJ. Dysmaturation of premature brain: importance, cellular mechanisms, and potential interventions. Pediatr Neurol. 2019;95:42-66.
7. Bonifacio SL, Van Meurs K. Neonatal neurocritical care: providing brain-focused care for all at risk neonates. Semin Pediatr Neurol. 2019;32:100774.
8. Tataranno ML, Vijlbrief DC, Dudink J, Benders MJNL. Precision medicine in neonates: a tailored approach to neonatal brain injury. Front Pediatr. 2021;9:634092.
9. El-Dib M, Abend NS, Austin T, Boylan G, Chock V, Mathur A, et al. Neuromonitoring in neonatal critical care part II: extremely premature infants and critically ill neonates. Pediatr Res. 2023;94(1):55-63.
10. Toet MC, Lemmers PMA. Brain monitoring in neonates. Early Hum Dev. 2009;85(2):77-84.
11. Iyer KK, Roberts JA, Hellström-Westas L, Wikström S, Hansen Pupp I, Ley D, et al. Cortical burst dynamics predict clinical outcome early in extremely preterm infants. Brain. 2015;138(8):2206-18.
12. Klebermass K, Olischar M, Waldhoer T, Fuiko R, Pollak A, Weninger M. Amplitude-integrated EEG pattern predicts further outcome in preterm infants. Pediatr Res. 2011;70(1):102-8.
13. van 't Westende C, Geraedts VJ, van Ramesdonk T, Benders MJNL, de Vries LS, van den Heuvel MP, et al. Neonatal quantitative electroencephalography and long-term outcomes: a systematic review. Dev Med Child Neurol. 2022;64(4):413-20.
14. Nordvik T, Schumacher EM, Larsson PG, Pripp AH, Løhaugen GC, Stiris T. Early spectral EEG in preterm infants correlates with neurocognitive outcomes in late childhood. Pediatr Res. 2022;92(4):1132-9.

15. Richardson J, Goshen S, Meledin I, Golan A, Goldstein E, Shany E. Predictive value of early amplitude integrated EEG in extremely premature infants. J Child Neurol. 2020;35(11):737-43.
16. Shibasaki J, Toyoshima K, Kishigami M. Blood pressure and aEEG in the 96h after birth and correlations with neurodevelopmental outcome in extremely preterm infants. Early Hum Dev. 2016;101:79-84.
17. Welch C, Helderman J, Williamson E, O'Shea TM. Brain wave maturation and neurodevelopmental outcome in extremely low gestational age neonates. J Perinatol. 2013;33(11):867-71.
18. Wikström S, Pupp IH, Rosén I, Normann E, Fellman V, Ley D, et al. Early single-channel aEEG/EEG predicts outcome in very preterm infants. Acta Paediatr. 2012;101(7):719-26.
19. Leen AT, Ouwehand S, Olsthoorn M, de Jonge RCJ, van der Grond J, Roest AAW, et al. Electroencephalography and brain magnetic resonance imaging in asphyxia comparing cooled and non-cooled infants. Eur J Paediatr Neurol. 2019;23(1):181-90.
20. Cainelli E, Vedovelli L, Wigley ILCM, Bisiacchi PS, Suppiej A. Neonatal spectral EEG is prognostic of cognitive abilities at school age in premature infants without overt brain damage. Eur J Pediatr. 2021;180(3):909-18.
21. Bourel-Ponchel E, Gueden S, Hasaerts D, Héberlé C, Mony L, Delignières A, et al. Normal EEG during the neonatal period: maturational aspects from premature to full-term newborns. Neurophysiol Clin. 2021;51(1):61-88.
22. Fogtmann EP, Plomgaard AM, Greisen G, Gluud C. Prognostic accuracy of electroencephalograms in preterm infants: a systematic review. Pediatrics. 2017;139(2):e20161951.
23. O'Toole JM, Boylan GB. Quantitative preterm EEG analysis: the need for caution in using modern data science techniques. Front Pediatr. 2019;7:174.
24. van Boven MR, Henke CE, Leemhuis AG, Pijpers I, de Vries LS, Oosterlaan J, et al. Machine learning prediction models for neurodevelopmental outcome after preterm birth: a scoping review and new machine learning evaluation framework. Pediatrics. 2022;150(3):e2021056052.
25. Hellström-Westas L, Rosén I, de Vries LS, Greisen G. Amplitude-integrated EEG classification and interpretation in preterm and term infants. Neoreviews. 2006;7(2):e76-87.
26. O'Toole JM, Boylan GB. NEURAL: quantitative features for newborn EEG using MATLAB. arXiv [Preprint]. 2017 Apr 19 [cited 2026 Aug 20]. doi:10.48550/arXiv.1704.05694
27. Bayley N. Bayley Scales of Infant and Toddler Development. 3rd ed. San Antonio (TX): Pearson; 2006.
28. Achenbach TM, Ruffle TM. The Child Behavior Checklist and related forms for assessing behavioral/emotional problems and competencies. Pediatr Rev. 2000;21(8):265-71.
29. Wechsler D. Wechsler Preschool & Primary Scale of Intelligence. 3rd ed. San Antonio (TX): The Psychological Corporation; 2002.
30. Henderson SE, Sugden D, Barnett AL. Movement Assessment Battery for Children-2. Washington (DC): APA PsycTests; 1992.