DeepCRI: Real-time EEG-based Prognostication after Cardiac Arrest

Michel J.A.M. van Putten[1,2], Hanneke M. Keijzer[3], Jeannette Hofmeijer[1,3], Pauline Buscher-Jungerhans[5], Nicolas Gaspard[6], Sarah Caroyer[6], Albertus Beishuizen[4], and Marleen C. Tjepkema-Cloostermans[1,2]

[1]University of Twente, Enschede, the Netherlands
[2]Medisch Spectrum Twente, Enschede, the Netherlands
[3]Rijnstate Hospital, Dept of Neurology, Arnhem, the Netherlands
[4]Medisch Spectrum Twente, Dept of Intensive Care Medicine, Enschede, the Netherlands
[5]Rijnstate Hospital, Dept of Intensive Care Medicine, Arnhem, the Netherlands
[6]Hôpital Universitaire de Bruxelles/Academisch Ziekenhuis Brussel, Dept of Intensive Care Medicine, Brussels, Belgium

Corresponding author:
Michel J.A.M. van Putten, *m.j.a.m.vanputten@utwente.nl*

Abstract

Accurate prediction of neurological outcome after cardiac arrest is essential for guiding intensive care decisions. Electroencephalography (EEG) supports prognostication; however, interpretation relies on expert judgment and is often subjective and delayed.

We developed DeepCRI, a bedside-integrated deep learning system that produces continuously updated prognostic trajectories during the first 36 hours after arrest. DeepCRI uses time-dependent decision boundaries to define good-, poor-, and gray-zone regions over time, and applies a lock-in rule that fixes classification only after sustained, concordant high-confidence evidence within a compact temporal window, thereby preventing transient threshold crossings from driving decisions.

During model development in a cohort of 522 patients, DeepCRI achieved an area under the receiver operating characteristic curve (AUC) of 0.97 at 24 h, with low calibration error (ECE=0.049). Independent validation was performed in an internal (n=219) and an external cohort (n=167). In the internal validation, DeepCRI provided lock-in classifications in 82.6% of patients within 24 h, achieving 100% specificity for poor outcome and 93.0% sensitivity for good outcome at 81.6% specificity; 17.4% remained in the gray zone. Performance in the external validation set was reduced: 62.3% locked within 24 h, and three false predictions of poor-outcome resulted in a poor-outcome specificity of 94.6%. Post hoc analysis indicated residual EMG artifacts contributed to these false poor-outcome predictions.

By embedding DeepCRI into routine ICU EEG infrastructure, we demonstrate the technical feasibility and clinical promise of continuous, real-time AI-driven prognostication for comatose patients after cardiac arrest.

1 Introduction

Reliable early prediction of neurological outcome after cardiac arrest helps ICU management, including family counseling and decisions on the continuation of care. Continuous EEG (cEEG) has emerged as a valuable tool allowing identification of patterns with high prognostic value, such as early return of physiological rhythms indicating a good outcome or burst-suppression with identical bursts indicating a poor outcome[1–8]. However, visual EEG interpretation is time-consuming, subjective, and limited by substantial interrater variability, even when standardized terminology is applied[9].

To address these limitations, several machine learning models—including deep neural networks—have been proposed to support EEG interpretation for prognostication[10–14]. While these approaches report sensitivities of 50-70% at high specificity, calibration is often neglected, and predictions are typically restricted to a single, fixed time point — usually 24 h post-arrest. Such static predictions fail to capture the dynamic evolution of post-anoxic encephalopathy in individual patients. Moreover, none of the previously proposed models has been technically implemented for bedside use, substantially limiting clinical translation.

Here, we introduce DeepCRI to overcome these barriers. DeepCRI is trained on a large, prospectively collected multicenter dataset, incorporates explicit calibration, and generates continuously updated prognostic estimates over the first 36 h after cardiac arrest. A dynamic lock-in mechanism ensures that once predictions become sufficiently stable, they are fixed and no longer subject to fluctuations. This enables individualized timing of prognostication and prevents oscillatory classification. In this study, we describe the development of DeepCRI, evaluate its performance in independent validation cohorts, and demonstrate its technical integration into routine ICU EEG monitoring.

2 Results

DeepCRI was trained on EEG recordings from 522 patients, of whom 51% had a good outcome. During model development (5-fold cross-validation for hyperparameter selection), DeepCRI achieved AUC=0.97 at 24 h with ECE=0.049, consistent with well-separated predicted-probability distributions (Figure 1-right panel).

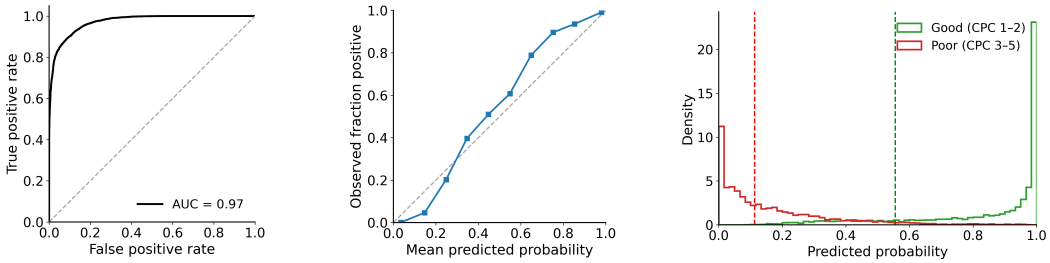

Figure 1: Left: ROC curve in the training cohort at t = 24 h (AUC = 0.97). Middle: calibration curve, with ECE = 0.049. Right: Predicted-probability distributions for good (green) and poor (red) outcomes. Dashed lines mark decision cut-offs chosen to achieve 100% specificity for "poor" and 95% specificity for "good", computed on the training set at t = 24 h.

To enable prediction at earlier time points, time-dependent decision boundaries were derived from the training set and fitted by exponential functions (See Supplementary Results: dynamic boundaries). These boundaries were applied unchanged in both validation cohorts (Figure 2; Table 1).

2.1 Lock-in parameter robustness

In the internal validation cohort, lock-in performance was robust across a range of parameter settings. Optimal performance required nine (poor) and thirteen (good) DeepCRI values occurring within a temporally compact window of 120 (poor) and 160 (good) minutes, with the minimum time required for poor-outcome prediction being 12 hours. Sensitivity and specificity were stable across parameter perturbations, whereas lock-in time showed only small shifts (few-percent range), indicating that classification was not dependent on fine-tuned lock-in settings. Once locked, classification remained unchanged. In a sensitivity analysis without automated artifact rejection, lock-in parameters shifted modestly, with slightly more false predictions of good outcome but no increase in false predictions of poor outcome (Supplementary Table 1).

2.2 Internal validation performance

Poor or good outcome classification in the internal validation cohort (n=219) was achieved in 181 patients (82.6%) within the first 24 hours after arrest. Sensitivity for good outcome was 93.0% (specificity 81.6%) and sensitivity for poor outcome was 59.0% (specificity 100%). Thirty-eight patients (17.4%) remained in the gray zone; these typically showed limited temporal EEG evolution. There were 14 false predictions of good outcome and no false predictions of poor outcome. Representative trajectories illustrating good lock-in, poor lock-in, and persistent gray-zone behaviour are shown in Figure 2.

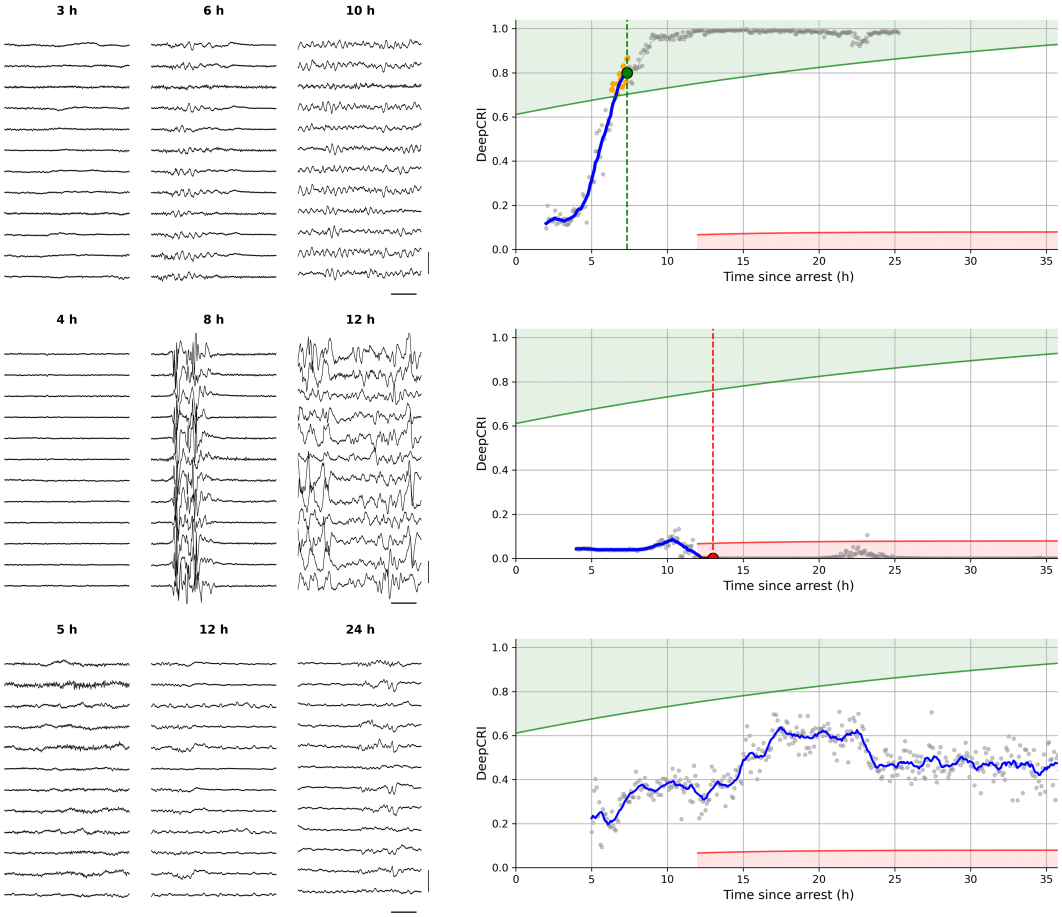

Figure 2: Five-second EEG epochs (left) with a reduced number of channels for display purposes and (right) corresponding DeepCRI trend curves for three patients. Times above EEG epochs indicate time since cardiac arrest. Top panel: Good outcome (CPC=1). Middle: Poor outcome (CPC=5). Bottom: DeepCRI in a patient where reliable prediction was not possible (eventually, this patient recovered well (CPC=1)). Good and poor outcome were predicted at lock-in times of 7 h and 12 h, indicated with the vertical dashed line. The dynamic boundaries for the good and poor regions are indicated with the solid green and red line, respectively. Vertical bar represents 100 µV, horizontal bar represents 1 s. EEG filter settings 1-25 Hz.

2.3 External validation performance

In the external cohort (n=167), overall performance was lower and 104 patients (62.3%) achieved lock-in within 24 h. Three false poor-outcome predictions occurred. Post hoc analysis indicated abundant EMG artifacts with incomplete removal by the artifact-rejection algorithm. EEG recordings also started later, and a smaller fraction of epochs survived artifact rejection (Supplementary Figure 1). Performance on the validation sets is summarized in Table 1.

Table 1: Performance of DeepCRI in the internal and external validation sets.

	Internal validation		**External validation**	
Metric	**Value**	**95% CI**	**Value**	**95% CI**
Total patients	219	–	167	–
Patients classified (locked)	181 (82.6%)	–	104 (62.3%)	–
Gray zone patients	38 (17.4%)	–	63 (37.7%)	–
For good outcome				
Sensitivity	93.0%	87.2–96.2%	69.7%	58.6–78.9%
Specificity	81.6%	71.4–88.7%	77.3%	65.8–85.7%
For poor outcome				
Sensitivity	59.0%	49.5–68.0%	46.8%	37.7–56.1%
Specificity	100.0%	97.0–100.0%	94.6%	85.4–98.1%
False predictions of good outcome	14	–	15	–
False predictions of poor outcome	0	–	3	–

2.4 Lock-in times and bedside integration

Median lock-in times ranged from 12-18 h (good outcome) and 15-19 h (poor outcome) across cohorts (Figure 3). DeepCRI has been integrated into the clinical EEG software (NeuroCenter EEG) and is currently being evaluated in the ICU (Supplementary video S1).

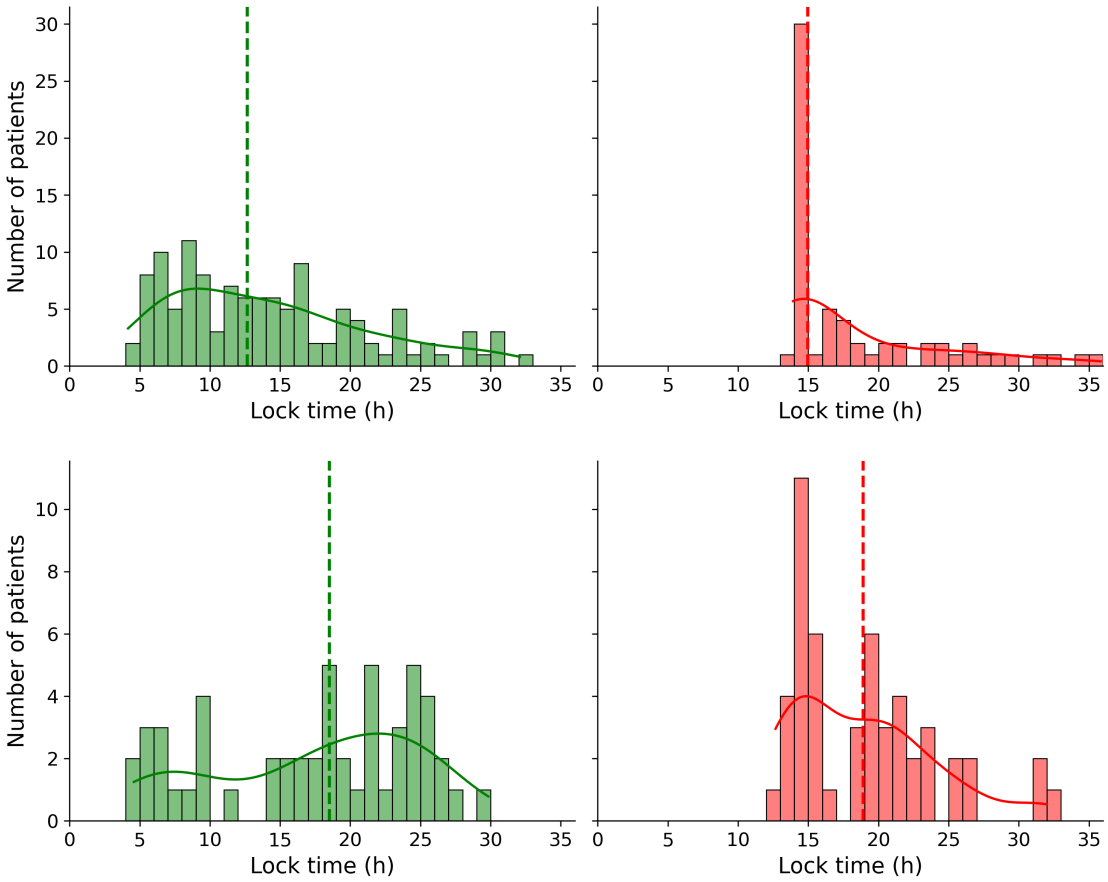

Figure 3: Histograms of lock-in times for patients with a good (left) and poor (right) outcome. Top: internal validation cohort. Bottom: external validation cohort. The smooth curves represent a kernel density estimate (KDE), highlighting the distribution shape and its mode. Dashed lines indicate the median lock-in times. In the internal validation, median lock-in times were 12 h for good and 15 h for poor outcome. In the external validation, these were 18 h and 19 h for good and poor outcome, respectively.

3 Discussion

DeepCRI is a calibrated deep learning model for EEG-based prognostication of neurological outcome in comatose patients after cardiac arrest, integrated directly into continuous bedside EEG. Unlike traditional approaches that provide static predictions, DeepCRI delivers dynamic predictions by tracking a continuously updated trajectory of EEG dynamics over time, enabling individualized timing of prognostication.

In the validation cohorts, DeepCRI yielded classifications for 62-83% of patients, achieving performance that equals or exceeds prior approaches[7;12;14–16], reporting sensitivities of 50–60% for poor outcome (at 99% specificity) and 60–75% for good outcome (at 95% specificity). External validation showed modestly reduced performance, including three false poor-outcome predictions attributable to residual EMG contamination, highlighting the need for robust artifact handling and expert visual EEG review in clinical settings. The higher gray-zone percentage in the external validation cohort likely reflected later EEG initiation and increased epoch rejections by the artifact detection algorithm.

Although the training dataset contained EEG recordings with EMG contamination and other artifacts, we employed an additional artifact-rejection algorithm based on traditional feature extraction methods. For improved EMG detection in future iterations, deep learning approaches, such as transformers or mixture-of-experts frameworks, could be explored, building on recent advances that demonstrate enhanced performance in artifact removal[17;18].

The lock-in mechanism is biologically motivated by the pathophysiology of postanoxic encephalopathy: transitions from a persistently poor EEG pattern to a sustained good pattern, or vice versa, are rare once a stable prognostic state has been reached. By requiring repeated, concordant high-confidence predictions within a limited temporal window, the lock-in mechanism aligns classification with this biological irreversibility. It reduces the risk of premature decisions driven by transient EEG fluctuations. Lock-in parameters were selected in a data-driven manner using the internal validation cohort and subsequently fixed, not adjusted during external validation. The model itself (training and per-epoch probability outputs) remained unchanged; only the post-processing decision rule affected lock-in timing.

Mean lock-in times were all within 24 hours, underscoring DeepCRI's potential for earlier, individualized ICU decision-making. Our findings also confirm prior evidence that early EEG contains strongly prognostic information[1;19] and that sedation does not substantially reduce its predictive value[20]. Recent work provides additional converging evidence that most EEG-based prognostic information is captured within the first 24 hours after cardiac arrest[21]. Early EEG (within 24 hours after cardiac arrest) apparently reflects the extent of irreversible neuronal injury more clearly than later recordings, when network reorganization can obscure the underlying pathology. For instance, highly pathological patterns such as identical burst suppression often evolve into periodic discharges that may eventually transition to a continuous background EEG[22], without clinical improvement, thereby reducing prognostic value over time.

DeepCRI has the potential to reduce reliance on EEG experts by providing a simple visual representation and quantitative interpretation at the bedside, enabling intensivists to derive prognostic information from the EEG with less dependence on neurologists. This is particularly valuable in hospitals where specialized neurophysiology resources may be limited, facilitating broader access to advanced prognostic information. DeepCRI is intended as an adjunct to—not a replace-

ment for—multimodal prognostication, aligning with international guidelines[23]. DeepCRI could integrate EEG with clinical examination, imaging, and biomarkers such as neuron-specific enolase (NSE) and neurofilament light (NfL)[24–27], potentially reducing uncertainty in family counseling and optimizing ICU resource allocation. Its early, high-confidence predictions may enable tailored clinical decisions, such as adjusting sedation protocols for patients with favorable EEG patterns. The ongoing SELECT trial (NCT06048796) is evaluating whether early EEG-informed individualized intensive care interventions can improve post-arrest outcomes.

Our study has limitations. As in all prognostication studies, the risk of self-fulfilling prophecy must be considered. In our cohorts, however, EEG was never used in isolation to guide treatment withdrawal. Since 2019, Dutch and Belgian guidelines recommend early EEG assessment after cardiac arrest, but always in combination with clinical examination and other modalities[28]. A poor outcome determination requires, among other factors, a persistently low Glasgow Coma Scale (GCS < 7) after complete cessation of sedatives, typically assessed at 36 hours or later. A second limitation is that, in the validation cohort, the outcome was defined retrospectively by CPC at hospital discharge rather than prospectively at 3–6 months. However, discharge CPC has been shown to correlate strongly with long-term outcome[29]. Finally, model explainability remains limited, though clinicians are presented with both the real-time EEG and the DeepCRI trend curve, enabling expert verification of the raw data and artifacts.

In conclusion, DeepCRI provides robust, real-time predictions of neurological outcomes in comatose patients within 12–24 h after arrest, while explicitly allowing abstention when reliable classification is not possible. Its dynamic confidence trajectories and lock-in mechanism reflect the evolving nature of post-anoxic encephalopathy and allow individualized timing of prognostication. In the external validation cohort, the few false-positive poor-outcome predictions were attributable to residual EMG contamination, underscoring the importance of robust artifact handling in real-world EEG recordings. Future work will focus on improving EMG detection and mitigation strategies. By integrating DeepCRI directly into bedside EEG software, we demonstrate the technical feasibility and clinical promise of AI-driven prognostication in intensive care medicine. Ongoing prospective studies will further assess clinical impact, usability, and ethical implications.

4 Methods

4.1 Patient population and outcome

The training data consisted of 522 cEEG recordings from ICU patients at two Dutch hospitals (2016–2022). Part of the EEG data was used in previous studies[11,30]. Outcome labels for the training data were the CPC scores at 3 or 6 months post-discharge, using the best (lowest) CPC values.

Two validation sets were used. An independent internal validation set comprising EEGs from 219 patients admitted during 2023-2024, obtained from MST and Rijnstate hospitals, and a second independent external validation set comprising 167 EEGs from Erasmus Hospital in Brussels, Belgium. EEG recordings were started within 24 h after cardiac arrest.

Patients were monitored using a 21-electrode EEG (10–20 system), starting within 12 to 24 hours post-arrest and continuing for 24 to 120 hours. The Medical Ethical Committee Twente approved the protocol (approval number K19-11). It waived the need for informed consent because

EEG monitoring and clinical follow-up are part of standard care at the participating centers. All ethical guidelines were followed.

Outcomes were dichotomized as either good (CPC 1–2; living independently with no or mild neurological deficits) or poor (CPC 3–5; severe neurological deficits, comatose, or brain death) neurological outcome. Labels for the validation sets were based on CPC at hospital discharge. Patients without follow-up or with non-neurological causes of death (e.g., hemodynamic failure or reinfarction) were excluded.

4.2 EEG preprocessing

Raw EEGs were recorded using the international 10–20 system at sampling rates of 500 or 512 Hz (MST and Rijnstate) and 250 or 256 Hz (Brussels). Signals were filtered with a 1–25 Hz FIR band-pass filter and downsampled to 64 Hz. Data were detrended and segmented into non-overlapping 10-second epochs. To ensure invariance to potential EEG polarity inversions (180° phase shifts), each epoch was analyzed in both its original and sign-reversed form, and predictions were averaged over the two versions.

All EEG epochs were automatically screened for artifacts using four complementary detectors: absolute amplitude (AMPabs), relative amplitude across channels (AMPrel), flat channels (FLT), and high-frequency contamination (FRQ; EMG/alpha ratio). Each 10-s segment was flagged if any channel exceeded predefined thresholds (AMPabs > 400 µV, AMPrel > 3, FLT > 0.8, FRQ > 0.3). Only windows with at least 15 artifact-free segments (out of 30) contributed to a DeepCRI value. If insufficient clean data were available, no DeepCRI was computed for that time window.

EEG data quality was quantified post hoc as the mean number of valid EEG segments retained after automated artifact rejection, with artifact rejection and quantification restricted to EEGs in which recording began within the first 24 h after cardiac arrest.

4.3 Model architecture

The CNN comprised 3 convolutional layers with ReLU activations, batch normalization after each, and spatial dropout (fixed at 0.2). This was followed by flattening, one dense layer (ReLU), dropout (0.3–0.7), and a sigmoid output layer for binary classification. Gaussian noise (0.05) was added to the input for robustness, with zero-padding before the first convolution and max pooling after the second and third convolutions. Kernel sizes were fixed at (3×3). Hyperparameters tuned via Optuna included convolutional filters (32–128 step 32 for layers 1 and 2; 64–256 step 64 for layer 3), dense units (64–512 step 64), L2 regularization (10^{-6} to 10^{-2}, log scale), and learning rates (10^{-6} to 10^{-3}, log scale). The Adam optimizer was used with early stopping (patience=10) and learning rate reduction on plateau (factor=0.2, patience=5). A custom loss function combined binary cross-entropy with tunable penalties for calibration error (weight 0.05–0.5 step 0.05) and separation of probability distributions (weight 0.1–1.0 step 0.1), plus a fixed specificity penalty (weight 1.0). The final model was retrained using the best hyperparameters found by Optuna. For deployment, the network was recompiled with standard binary cross-entropy loss to allow inference without custom loss definitions.

4.4 Model training and optimization

Model training employed 5-fold stratified group cross-validation on EEG data recorded 24 hours after cardiac arrest. Hyperparameters were optimized with Optuna[31] to maximize the area under the ROC curve (AUC) while minimizing calibration error (ECE), as described in[32], using a custom loss function incorporating specificity and separation penalties. The custom loss function combined binary cross-entropy for classification with an ECE term to ensure well-calibrated probabilities, plus penalties to enforce high specificity (e.g., minimizing false positives for poor outcomes) and maximize separation between good/poor outcome distributions. This optimization prioritized clinically relevant thresholds, such as 100% specificity for poor predictions.

4.5 Dynamic boundaries

To generate time-evolving predictions, we applied the trained model to additional time points within the first 36 h after cardiac arrest, using data from the training set. This yielded dynamic boundaries corresponding to the 95% and 100% specificity thresholds for good and poor outcomes, respectively. Three prognostic regions were defined: good outcome, poor outcome, and a 'gray zone' in between. Threshold values per hour were subsequently fitted with an exponential function, $a + b\exp(-t/\tau)$.

4.6 Classification and lock-in

For each patient, the classifier outputs a DeepCRI value for consecutive 5-minute windows. Each value represents the median prediction score derived from all artifact-free 10-s segments within that window, including scores obtained after signal inversion. A minimum of 15 artifact-free segments per window was required for inclusion.

Dynamic locking was introduced by defining a discrete stability criterion that requires repeated, concordant predictions within a limited temporal window. This approach prevents premature classification due to short-lived fluctuations around decision boundaries. It ensures that outcome classification reflects sustained and internally consistent prognostic evidence rather than transient threshold crossings. Once a lock is achieved, the classification is considered stable and irreversible.

Locking parameters were selected using the internal validation cohort, which was entirely independent of the model training data. A data-driven optimization procedure was applied using Optuna to explore a constrained set of five post-processing parameters governing (i) the minimum number of DeepCRI values required for lock-in, (ii) the temporal compactness of these values, and (iii) outcome-specific timing constraints. Specifically, the optimization explored plausible ranges for the minimum required number of DeepCRI values (typically 4–15), the duration of the temporal aggregation window (20 minutes to 3 hours), and the earliest permissible time for poor-outcome lock-in following return of spontaneous circulation. Optimization aimed to maximize sensitivity for outcome classification while enforcing a strict specificity constraint of 100% for poor-outcome prediction. Following optimization, we selected a representative parameter configuration from the stable performance region; this configuration was fixed and applied to the external validation cohort without further modification.

When a lock was identified, patients were classified as locked-good or locked-poor. Patients without a qualifying sequence remained in the gray zone. No temporal smoothing was applied to

DeepCRI values for classification.

4.7 Model performance and implementation

Performance was assessed using ROC AUC, sensitivity, specificity, calibration curves, and gray-zone proportion. Sensitivity and specificity confidence intervals were calculated using the Beta distribution, which provides more accurate coverage than the traditional Wald method, particularly when estimated proportions are close to zero or one[33]. Calibration, expressed as the expected calibration error (ECE), was evaluated only for fixed-time (24 h) probabilistic outputs, as these metrics do not apply to dynamic lock-in predictions.

The model was implemented in Python using TensorFlow/Keras with mixed precision training. GPU training was performed on an NVIDIA GTX 1080 under Rocky Linux 9. Training required approximately 5 days. The final network was deployed in the clinical EEG platform (NeuroCenter EEG, Clinical Science Systems, Leiden, the Netherlands), allowing real-time bedside inference in the ICU.

Author contributions

MvP: conceptualisation, coding, data analysis, writing first draft, and data collection; MT: conceptualization, coding, data analysis, and data collection; HK: data collection; NG: data collection; AB: data collection; JH: data collection; PB: data collection; All authors: critical reading and commenting on the first draft. All authors reviewed the manuscript.

Competing interest

MvP is co-founder of Clinical Science Systems, a manufacturer of EEG software (NeuroCenter EEG)

Acknowledgements

We thank the lab technicians of the participating centers for their assistance in recording the EEGs. Clinical Science Systems is acknowledged for creating the videos.

Funding

Not applicable.

Ethics statement

This study was approved by the Medical Ethical Committee Twente (approval number K19-11), and all ethical guidelines were followed.

References

[1] Ruijter, B. *et al.* Early electroencephalography for outcome prediction of postanoxic coma: A prospective cohort study. *Annals of Neurology* **86** (2019).

[2] Rossetti, A. O., Rabinstein, A. A. & Oddo, M. Neurological prognostication of outcome in patients in coma after cardiac arrest. *The Lancet Neurology* **15**, 597–609 (2016).

[3] Hockaday, J. M., Potts, F., Epstein, E., Bonazzi, A. & Schwab, R. S. Electroencephalographic changes in acute cerebral anoxia from cardiac or respiratory arrest. *Electroencephalogr Clin Neurophysiol* **18**, 575–586 (1965).

[4] Cloostermans, M., Meulen, F. V., Eertman, C., Hom, H. & Putten, M. V. Continuous electroencephalography monitoring for early prediction of neurological outcome in postanoxic patients after cardiac arrest: A prospective cohort study. *Critical Care Medicine* **40** (2012).

[5] Hofmeijer, J., Tjepkema-Cloostermans, M. C. & van Putten, M. J. Burst-suppression with identical bursts: a distinct eeg pattern with poor outcome in postanoxic coma. *Clinical Neurophysiology* (2013).

[6] van Putten, M. & van Putten, M. Uncommon eeg burst-suppression in severe postanoxic encephalopathy. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology* **121**, 1213–9 (2010).

[7] Westhall, E. *et al.* Standardized eeg interpretation accurately predicts prognosis after cardiac arrest. *Neurology* **86**, 1482–1490 (2016).

[8] Bronder, J., Cho, S.-M., Geocadin, R. G. & Ritzl, E. K. Revisiting eeg as part of the multidisciplinary approach to post-cardiac arrest care and prognostication: A review. *Resuscitation Plus* **9**, 100189 (2022).

[9] Hirsch, L. *et al.* American clinical neurophysiology society's standardized critical care eeg terminology: 2021 version. *Journal of clinical neurophysiology* **38**, 1–29 (2021).

[10] Tjepkema-Cloostermans, M. C., van Meulen, F. B., Meinsma, G. & van Putten, M. J. A. M. A cerebral recovery index (cri) for early prognosis in patients after cardiac arrest. *Crit Care* **17**, R252 (2013).

[11] Pham, S. D. T. *et al.* Outcome prediction of postanoxic coma: A comparison of automated electroencephalography analysis methods. *Neurocritical Care* (2022).

[12] Zheng, W. L. *et al.* Predicting neurological outcome in comatose patients after cardiac arrest with multiscale deep neural networks. *Resuscitation* **169**, 86–94 (2021).

[13] Zheng, W. L. *et al.* Predicting neurological outcome from electroencephalogram dynamics in comatose patients after cardiac arrest with deep learning. *IEEE Transactions on Biomedical Engineering* **69**, 1813–1825 (2022).

[14] Chen, C. C. *et al.* Electroencephalogram-based machine learning models to predict neurologic outcome after cardiac arrest: A systematic review. *Resuscitation* **194** (2024).

[15] Tjepkema-Cloostermans, M. C. *et al.* Outcome prediction in postanoxic coma with deep learning. *Critical Care Medicine* 1 (2019).

[16] Zubler, F. & Tzovara, A. Deep learning for eeg-based prognostication after cardiac arrest: from current research to future clinical applications. *Frontiers in Neurology* **14** (2023).

[17] Raj, V. A., Parupudi, T., Thalengala, A. & Nayak, S. G. A comprehensive review of deep learning models for denoising eeg signals: challenges, advances, and future directions. *Discover Applied Sciences* **7**, 1–32 (2025).

[18] Choi, B. J., Milsap, G., Scholl, C. A., Tenore, F. & Ogg, M. A statistical mixture-of-experts framework for emg artifact removal in eeg: Empirical insights and a proof-of-concept application. *https://arxiv.org/abs/2509.19385* (2025).

[19] Admiraal, M. *et al.* Assessing both early and late eeg patterns improves prediction of outcome after cardiac arrest. *Resuscitation* 110762 (2025).

[20] Ruijter, B. J., van Putten, M. J., van den Bergh, W. M., Tromp, S. C. & Hofmeijer, J. Propofol does not affect the reliability of early eeg for outcome prediction of comatose patients after cardiac arrest. *Clinical Neurophysiology* **130**, 1263–1270 (2019).

[21] Hodžić, I., Doelkahar, B., Horn, J. & van Rootselaar, A.-F. A continuous convolutional neural network: very early eeg most predictive for poor neurological outcome in postanoxic coma. *Resuscitation* 110943 (2025).

[22] Westhall, E. *et al.* Time to epileptiform activity and eeg background recovery are independent predictors after cardiac arrest. *Clinical Neurophysiology* **129**, 1660–1668 (2018).

[23] Nolan, J. P. *et al.* European resuscitation council and european society of intensive care medicine guidelines 2025: post-resuscitation care. *Intensive Care Medicine* (2025).

[24] Disanto, G. *et al.* Longitudinal serum neurofilament light kinetics in post- anoxic encephalopathy. *Annals of Clinical and Translational Neurology* **10**, 2407–2412 (2023).

[25] Meyer, M. A. S. *et al.* Neurofilament light chain for prognostication after cardiac arrest-first steps towards validation. *Critical Care* **29**, 348 (2025).

[26] Strate, I. A. *et al.* Early eeg and serum biomarkers for prognostication after cardiac arrest. *Resuscitation Plus* 101154 (2025).

[27] Moseby-Knappe, M. *et al.* Blood biomarkers for the prediction of outcome after cardiac arrest: an international prospective observational study within the targeted hypothermia versus normothermia after out-of-hospital cardiac arrest (ttm2) trial. *The Lancet Respiratory Medicine* (2025).

[28] Taccone, F. S. *et al.* Neuroprognostication after adult cardiac arrest treated with targeted temperature management: task force for belgian recommendations. *Acta Neurologica Belgica* **117**, 3–15 (2017).

[29] Wu, W., Chopra, A., Ziegler, C., McLeod, S. L. & Lin, S. Predictive value of hospital discharge neurological outcome scores for long-term neurological status following out-of-hospital cardiac arrest: A systematic review. *Resuscitation* **151**, 139–144 (2020).

[30] Tjepkema-Cloostermans, M. *et al.* Outcome prediction in postanoxic coma with deep learning. *Critical care medicine* **47** (2019).

[31] Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631 (ACM, 2019).

[32] Guo, C., Pleiss, G., Sun, Y. & Weinberger, K. Q. On calibration of modern neural networks. In *Proceedings of the 34 th International Conference on Machine Learning* (PMLR 70, 2017).

[33] Brown, L. D., Cai, T. T. & Dasgupta, A. Interval estimation for a binomial proportion. *Statistical Science* **16**, 101–133 (2001).