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

Michel J.A.M. van Putten[a,b], Hanneke M. Keijzer[c], Jeannette Hofmeijer[a,c], Pauline Buscher-Jungerhans[d], Albertus Beishuizen[e], Marleen C. Tjepkema-Cloostermans[a,b]

[a]*University of Twente, Enschede, the Netherlands*
[b]*Medisch Spectrum Twente, Koningsplein 1, Enschede, the Netherlands*
[c]*Rijnstate Hospital, dept of neurology, Arnhem, the Netherlands*
[d]*Rijnstate Hospital, dept of Intensive Care Medicine, Arnhem, the Netherlands*
[e]*Medisch Spectrum Twente, Dept of Intensive Care Medicine, Enschede, the Netherlands*

Abstract

Rationale Accurate early prediction of neurological outcome after cardiac arrest is essential for guiding intensive care decisions. Visual EEG interpretation can support prognostication but is time-consuming and subjective.

Methods We developed DeepCRI, a deep learning model trained on EEG recordings obtained 24 h after arrest from a prospective cohort of 522 patients across two centers. Labels were based on Cerebral Performance Category (CPC) scores, restricted to neurologically determined outcomes. Calibration was assessed using the expected calibration error (ECE). To enable dynamic prognostication, we derived boundaries for the 95% and 100% specificity thresholds (good and poor outcome, respectively) by applying the trained model to the training cohort at all time points between 0–36 h after arrest. Validation was performed in an independent cohort of 222 patients (60% good outcome), after automatic artifact rejection. Predictions were locked once DeepCRI values consistently exceeded the predefined boundaries for good or poor outcome. DeepCRI was implemented in Neuro- Center EEG for real-time bedside monitoring.

Results In training, area under the receiver operating characteristic curve (ROC AUC) at 24 h was 0.97 with ECE = 0.049. In the validation cohort, outcome classification was achieved in 184/222 patients (83%) within 24 h; 38 patients (17%) remained in the gray zone. Sensitivity for good out- come was 93.9% (95% confidence interval [CI]: 88.5–96.9%) at 78.9% (95% CI: 67.1–84.6%) specificity. Sensitivity for poor outcome was 56.6% (95% CI: 47.1–65.7%) at 100% (95% CI: 97.1–100%) specificity. ROC AUC at 24 h was 0.91 (95% CI: 0.87–0.94). Median lock-in time was 11.2 h for good and 14.4 h for poor outcome.

Conclusion DeepCRI is the first real-time deep learning model for prognostication of neurological outcome within 12-24 h after cardiac arrest. Its lock-in mechanism ensures individualized, stable, classifications. By embedding DeepCRI into routine EEG infrastructure, we demonstrate the technical feasibility of continuous bedside integration 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 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 for good outcome or burst-suppression with identical bursts for poor outcome [1, 2, 3, 4, 5, 6, 7, 8]. Yet, visual EEG interpretation is time-consuming, subjective, and limited by interrater variability—even with standardized terminology [9].

Several machine learning models— including deep neural networks—have been proposed to support in EEG interpretation prognostication [10, 11, 12, 13, 14], reporting sensitivities of 50-70% at high specificity. However, calibration was often neglected and predictions were typically restricted to a single, fixed time point — often 24 h post-arrest — which does not reflect the dynamic evolution of post-anoxic encephalopathy in individual patients. Moreover, none of the previously proposed models reached technical implementation for bedside use. These limitations have hampered clinical translation.

Here we introduce DeepCRI to overcome these barriers. DeepCRI is trained on a large, prospectively collected multicenter dataset, explicitly incorporates calibration and generates continuously updated prognostic estimates over the first 36 h after cardiac arrest. A lock-in mechanism ensures that once predictions are sufficiently stable, they are fixed and no longer subject to fluctuation, thereby allowing individualized timing of prognostication and preventing oscillating or premature classifications. In this study, we present the development of DeepCRI, evaluate its performance in an independent validation cohort, and demonstrate its technical integration into routine ICU EEG monitoring.

2. Results

DeepCRI was trained on EEG recordings of 522 patients, of which 51% had a good outcome. The ROC (at t=24 h), expected calibration error (ECE), and separation of patients with a good and poor outcome as obtained from the training are shown in Figure 1.

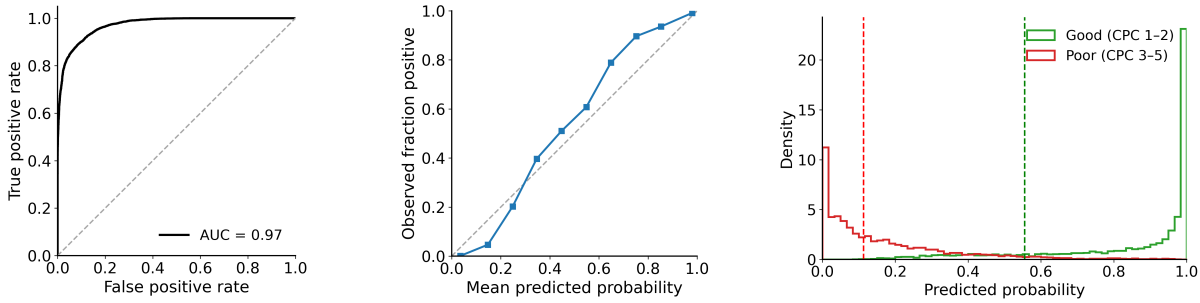

Figure 1: Left: ROC curve on the training set at t = 24 h (AUC = 0.97). Middle: calibration curve, with ECE = 0.049. Right: Predicted-probability distributions for good (green) and poor (red) outcomes. Dashed lines mark decision cut-offs chosen to achieve 100% specificity for "poor" and 95% specificity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

At 24 h in the training set, the AUC=0.97 of the ROC with ECE = 0.049, in agreement with the well-separated predicted probability distributions (Fig. 1-right panel). Performance on the independent validation set of 222 recordings (60% good outcome) is summarized in Table 1.

Table 1: Performance of DeepCRI in the external validation cohort (n=222). Outcomes were locked after ≥ 8 non-gray zone predictions within 60 min; poor-outcome locks were restricted to ≥ 12 h post-arrest. Traditional ROC analysis at 24 h is included for comparison.

Metric	**Value**	**95% CI**
Total patients	222	–
Patients classified (locked)	184 (83%)	–
Gray zone patients	38 (17%)	–
For good outcome		
Sensitivity	93.9%	88.5–96.9%
Specificity	78.9%	68.5–86.6%
For poor outcome		
Sensitivity	56.6%	47.1–65.7%
Specificity	100%	97.1–100%
False predictions of good outcome	16	–
False predictions of poor outcome	0	–
ROC AUC at 24 h (non-locking)	0.91	0.87–0.94

Outcome classification was achieved in 184 of 222 patients (83%) within the first 24 hours after arrest. The sensitivity for the prediction of a good outcome was 93.9% at a specificity of 78.9%. The sensitivity for the prediction of a poor outcome was 56.6% at a specificity of 100%. A total

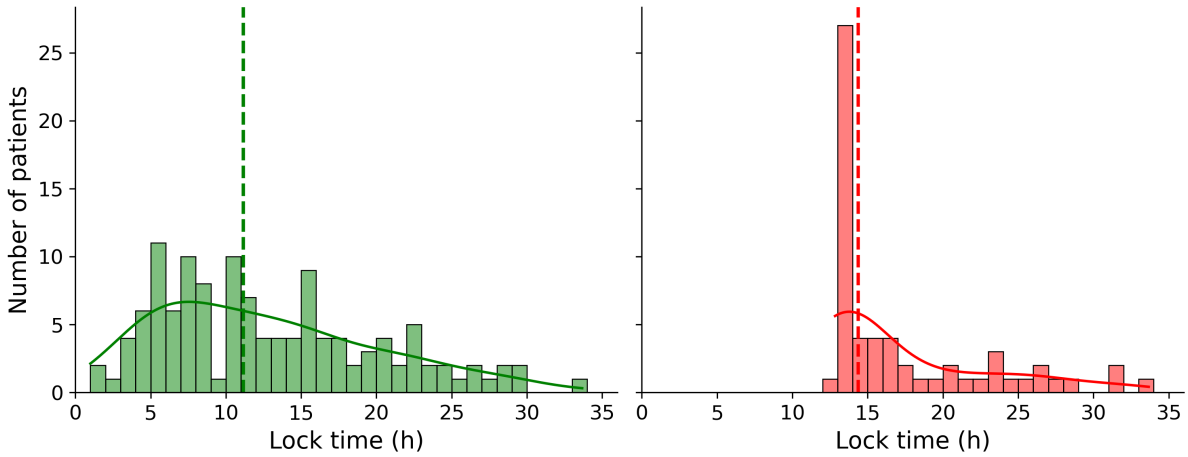

Figure 2: Histogram of lock times for patients with a good (left) and poor (right) outcome in the validation cohort. The algorithm did not classify patients with a poor outcome earlier than 12 h after arrest, setting the lower limit. The smooth curve represents a kernel density estimate (KDE), highlighting the distribution shape and its mode. Dashed lines indicate the median lock-in time, with 11.2 h for patients with a good and 14.4 h for patients with a poor outcome.

of 38 patients (17%) remained in the gray zone. Of these patients, 30 had a poor outcome. The median lock-in time was 11.2 h for patients with a good outcome and 14.4 h for those with a poor outcome, where lock-in time is defined as the point when predictions stabilize after a sequence of consistent high-confidence DeepCRI values. The distribution of lock-in times in the validation cohort is shown in Figure 2.

Once locked, outcome predictions remained unchanged. Gray zone classifications typically occurred in patients with minimal temporal EEG evolution. Figure 3 shows representative cases with different outcome trajectories. There were 16 false predictions of good outcome and no false predictions of poor outcome. We repeated the analysis without artifact rejection. Performance metrics were slightly higher, but at the cost of one false poor prediction (Supplement 2).

DeepCRI has been integrated into the clinical EEG software (NeuroCenter EEG) and is currently evaluated in the ICU setting. Real-time evolution of predictions using DeepCRI in two patients is shown in Supplementary Videos S1.

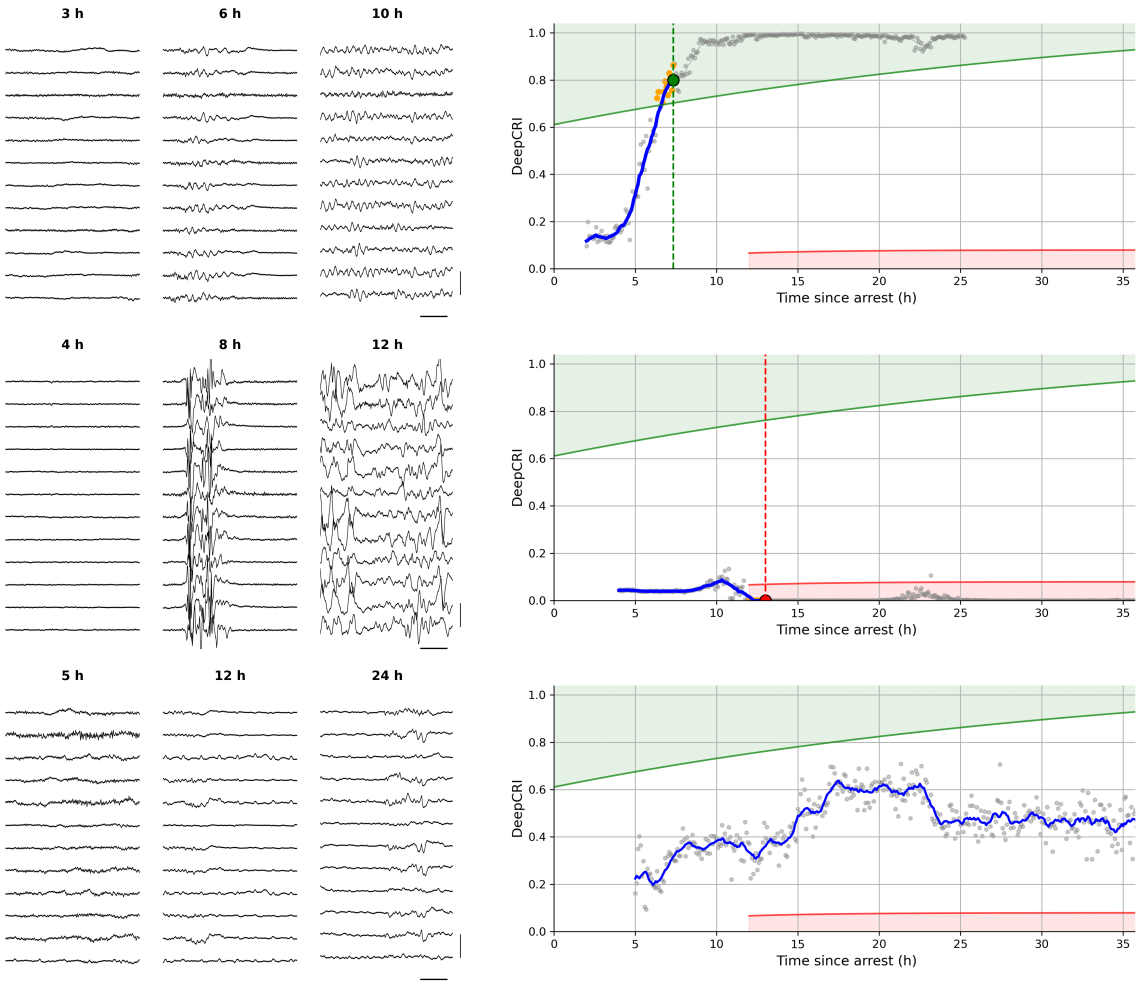

Figure 3: Five second EEG epochs (left) with a reduced number of channels for display purposes and (right) corresponding DeepCRI trend curves for three patients. Times above EEG epochs indicate time since cardiac arrest. Top panel: Good outcome (CPC=1). Middle: Poor outcome (CPC=5). Bottom: DeepCRI in 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.

3. Discussion

DeepCRI is the first real-time, calibrated deep learning model for neurological prognostication of comatose cardiac arrest patients, integrated directly into continuous EEG at the bedside. DeepCRI provides predictions that are comparable to or exceed visual EEG interpretation. Unlike traditional models that rely on static predictions at fixed time points, DeepCRI is the first to deliver dynamic predictions, through a continuously updating trajectory of the EEG dynamics over time. This enables individualized timing of outcome prediction. Trained on 24-hour data, DeepCRI generalizes well to earlier time points with adjusted cut-off values. The lock-in mechanism is biologically motivated, as transitions from a persistently poor or good EEG are very unlikely. Therefore, once DeepCRI shows sustained high-confidence predictions, classification becomes fixed.

In validation, DeepCRI provided confident predictions in 83% of patients, with mean lock-in times of 11 h for good and 14 h for poor outcome. Herewith, DeepCRI achieves performance that equals or exceeds prior approaches [15, 7, 12, 16, 14], reporting sensitivities of 50–60% for poor outcome (at 99% specificity) and 60–75% for good outcome (at 95% specificity). In our primary artifact-rejected analysis, specificity for poor outcome reached 100%, minimizing false pessimistic predictions. Patients who remained in the gray zone (17%) typically showed limited temporal EEG evolution, consistent with the model's principle to withhold classification in the absence of clear prognostic signals. Automatic artifact rejection modestly reduced the proportion of classified patients and the sensitivity for poor outcome, while eliminating false poor predictions. Without artifact rejection, sensitivity for poor outcome was somewhat higher, but at the expense of one false poor prediction (see Supplementary Table).

Our findings confirm prior evidence that early EEG contains strongly prognostic information [1, 17] and that sedation does not substantially reduce its predictive value [18]. 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 and, eventually, a continuous background [19], without clinical improvements, reducing prognostic values over time.

An important asset of DeepCRI is its 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 replacement for—multimodal prognostication, aligning with international guidelines [20]. DeepCRI integrates EEG with clinical examination, imaging, and biomarkers like Neuron Specific Enolase (NSE) or Neurofilament light (NfL) [21, 22], potentially reducing uncertainty in family counseling and optimizing ICU resource allocation. Its early, high-confidence predictions (83% classified within 24 h) 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.

This 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 withdrawal of treatment. Since 2019, Dutch guidelines recommend early EEG assessment after cardiac arrest, but always in combination with clinical examination and other modalities. 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 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 [23]. Third, our cohorts are from Dutch hospitals; validation in more diverse populations is needed to ensure generalizability. 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 outcome prediction of comatose patients within 12 to 24 h after arrest. Its dynamic confidence trajectories and biologically grounded lock-in mechanism reflect the evolving nature of post-anoxic encephalopathy and allow individualized timing of prediction. Importantly, we demonstrate the first technical integration of a deep learning prognostic model directly into continuous EEG software at the bedside. Formal prospective evaluation of clinical impact, usability, and ethical implications is ongoing.

4. Methods

B.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 [24, 11]. An independent validation set of 222 patients from 2023–2024 was used for testing. All patients were monitored using 21-electrode EEG (10–20 system), starting within 12 to 24 hours post-arrest and continuing for 24 to 120 hours. Recordings were filtered (1–25 Hz), downsampled to 64 Hz, and segmented into 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. The Medical Ethical Committee Twente approved the protocol and waived the need for informed consent as EEG monitoring and clinical follow-up are part of standard care in the participating centers.

Artifact rejection was performed automatically using a custom-made algorithm. Segments contaminated by muscle, movement, or electrode artifacts were excluded from model input. The procedure was applied uniformly to all EEGs prior to DeepCRI analysis, ensuring comparability across patients. For sensitivity analyses, we also performed the full analysis on the validation set without artifact rejection.

For the training data, cerebral performance category (CPC) scores were assessed at 3 or 6 months after arrest by telephone interview. For the independent validation, in several patients, we used electronic patient records to assess the CPC scores at 3 or 6 months; patients with missing scores or non-neurological causes of death were excluded. 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. More details are provided in Supplement A.1-A.2.

B.2. Model training and optimization

A convolutional neural network was trained using 5-fold stratified group cross-validation with EEG data recorded 24 hours after cardiac arrest. Hyperparameters were optimized with Optuna [25] to maximize the area under the ROC curve (AUC) while reducing calibration error (ECE), using a custom loss function enriched with 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. More details on the architecture (e.g., convolutional layers, kernel sizes) and tuning (e.g., learning rate, batch size) are in Supplement A.3-A.4.

To create predictions evolving over time, we applied our model to other time points within the first 36 h after the arrest of the training set, thus creating dynamic boundaries for the 95% and 100% specificity thresholds for good and poor outcomes, respectively. These two boundaries resulted in three prognostic regions: predicted good outcome, predicted poor outcome, and a 'gray zone', containing predictions that were between the good and poor boundaries. Threshold values per hour were subsequently fitted with two exponential functions (Supplement A.5-A.6 provides more details).

B.3. Dynamic classification and lock-in

For each patient, the classifier outputs the Cerebral Recovery Index (CRI) for subsequent 5 minute windows. Each CRI value is the median score from all 10 s segments in that window, including scores after signal inversion. At least 15 artefact-free segments per window are required. Each CRI value was classified relative to these curves with zero tolerance. The "locking" algorithm searched for sequences of at least eight CRI-values (within a period of 60 min) all lying in the same outcome zone. A "good" lock was permitted at any time after return of spontaneous circulation (ROSC). A "poor" lock was only accepted if the lock endpoint occurred ≥ 12.0 h post-arrest, to avoid premature pessimistic classification. The 'lock-in time' refers to the timestamp of the final point in this qualifying sequence, marking when the prediction stabilizes and becomes fixed (i.e., 'locked in').

When a lock is found, the patient is classified as "locked: good" or "locked: poor". Patients without a qualifying sequence remain in the gray zone. No smoothing is applied to CRI values for classification; moving average filtering (window = 12 points) was used only for plotting.

B.4. Model performance and implementation

Performance was assessed using ROC AUC, sensitivity, specificity, calibration curves, and gray-zone proportion. Calibration, expressed as the expected calibration error (ECE), was evaluated only for fixed-time (24 h) probabilistic outputs, as these metrics are not applicable to dynamic lock-in predictions. Confidence intervals were estimated using the beta-distribution method.

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.

References

[1] B.J. Ruijter, M.C. Tjepkema-Cloostermans, S.C. Tromp, W.M. van den Bergh, N.A. Foudraine, F.H.M. Kornips, G. Drost, E. Scholten, F.H. Bosch, A. Beishuizen, M.J.A.M. van Putten, and J. Hofmeijer. Early electroencephalography for outcome prediction of postanoxic coma: A prospective cohort study. *Annals of Neurology*, 86, 2019.

[2] Andrea O Rossetti, Alejandro A Rabinstein, and Mauro Oddo. Neurological prognostication of outcome in patients in coma after cardiac arrest. *The Lancet Neurology*, 15:597–609, 2016.

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

[4] M.C. Cloostermans, F.B. Van Meulen, C.J. Eertman, H.W. Hom, and M.J.A.M. Van Putten. 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] Jeannette Hofmeijer, Marleen C. Tjepkema-Cloostermans, and Michel J.a.M. van Putten. Burst-suppression with identical bursts: a distinct eeg pattern with poor outcome in postanoxic coma. *Clinical Neurophysiology*, 10 2013.

[6] M.J.A.M. Van Putten and J. Hofmeijer. Eeg monitoring in cerebral ischemia: Basic concepts and clinical applications. *Journal of Clinical Neurophysiology*, 33, 2016.

[7] Erik Westhall, Andrea O. Rossetti, Anne Fleur Van Rootselaar, Troels Wesenberg Kjaer, Janneke Horn, Susann Ullén, Hans Friberg, Niklas Nielsen, Ingmar Rosén, Anders Aneman, David Erlinge, Yvan Gasche, Christian Hassager, Jan Hovdenes, Jesper Kjaergaard, Michael Kuiper, Tommaso Pellis, Pascal Stammet, Michael Wanscher, Jørn Wetterslev, Matt P. Wise, and Tobias Cronberg. Standardized eeg interpretation accurately predicts prognosis after cardiac arrest. *Neurology*, 86:1482–1490, 2016.

[8] Jay Bronder, Sung-Min Cho, Romergryko G. Geocadin, and Eva Katharina Ritzl. Revisiting eeg as part of the multidisciplinary approach to post-cardiac arrest care and prognostication: A review. *Resuscitation Plus*, 9:100189, 2022.

[9] L.J. Hirsch, M.W.K. Fong, M. Leitinger, S.M. LaRoche, S. Beniczky, N.S. Abend, J.W. Lee, C.J. Wusthoff, C.D. Hahn, M.B. Westover, E.E. Gerard, S.T. Herman, H.A. Haider, G. Osman, A. Rodriguez-Ruiz, C.B. Maciel, E.J. Gilmore, A. Fernandez, E.S. Rosenthal, J. Claassen, A.M. Husain, J.Y. Yoo, E.L. So, P.W. Kaplan, M.R. Nuwer, M. van Putten, R. Sutter, F.W. Drislane, E. Trinka, and N. Gaspard. American clinical neurophysiology society's standardized critical care eeg terminology: 2021 version. *Journal of clinical neurophysiology*, 38:1–29, 2021.

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

[11] Stanley D. T. Pham, Hanneke M. Keijzer, Barry J. Ruijter, Antje A. Seeber, Erik Scholten, Gea Drost, Walter M. van den Bergh, Francois H. M. Kornips, Norbert A. Foudraine, Albertus

Beishuizen, Michiel J. Blans, Jeannette Hofmeijer, Michel J. A. M. van Putten, and Marleen C. Tjepkema-Cloostermans. Outcome prediction of postanoxic coma: A comparison of automated electroencephalography analysis methods. *Neurocritical Care*, 2022.

[12] Wei Long Zheng, Edilberto Amorim, Jin Jing, Wendong Ge, Shenda Hong, Ona Wu, Mohammad Ghassemi, Jong Woo Lee, Adithya Sivaraju, Trudy Pang, Susan T. Herman, Nicolas Gaspard, Barry J. Ruijter, Jimeng Sun, Marleen C. Tjepkema-Cloostermans, Jeannette Hofmeijer, Michel J.A.M. van Putten, and M. Brandon Westover. Predicting neurological outcome in comatose patients after cardiac arrest with multiscale deep neural networks. *Resuscitation*, 169:86–94, 2021.

[13] Wei Long Zheng, Edilberto Amorim, Jin Jing, Ona Wu, Mohammad Ghassemi, Jong Woo Lee, Adithya Sivaraju, Trudy Pang, Susan T. Herman, Nicolas Gaspard, Barry J. Ruijter, Marleen C. Tjepkema-Cloostermans, Jeannette Hofmeijer, Michel J.A.M. Van Putten, and M. Brandon Westover. 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] Chao Chen Chen, Shavonne L. Massey, Matthew P. Kirschen, Ian Yuan, Asif Padiyath, Allan F. Simpao, and Fuchiang Rich Tsui. Electroencephalogram-based machine learning models to predict neurologic outcome after cardiac arrest: A systematic review. *Resuscitation*, 194, 2024.

[15] Marleen C. Tjepkema-Cloostermans, Catarina da Silva Lourenço, Barry J. Ruijter, Selma C. Tromp, Gea Drost, Francois H. M. Kornips, Albertus Beishuizen, Frank H. Bosch, Jeannette Hofmeijer, and Michel J. A. M. van Putten. Outcome prediction in postanoxic coma with deep learning. *Critical Care Medicine*, page 1, 6 2019.

[16] Frederic Zubler and Athina Tzovara. Deep learning for eeg-based prognostication after cardiac arrest: from current research to future clinical applications. *Frontiers in Neurology*, 14, 2023.

[17] Marjolein Admiraal, Sofia Backman, Martin Annborn, Ola Borgquist, Josef Dankiewicz, Joachim Düring, Marion Moseby-Knappe, Stéphane Legriel, Hans Lindehammar, Anna Lybeck, Niklas Nielsen, Andrea O. Rossetti, Johan Undén, Tobias Cronberg, and Erik Westhall. Assessing both early and late eeg patterns improves prediction of outcome after cardiac arrest. *Resuscitation*, page 110762, 8 2025.

[18] Barry J. Ruijter, Michel J.A.M. van Putten, Walter M. van den Bergh, Selma C. Tromp, and Jeannette Hofmeijer. Propofol does not affect the reliability of early eeg for outcome prediction of comatose patients after cardiac arrest. *Clinical Neurophysiology*, 130:1263–1270, 2019.

[19] E. Westhall, I. Rosén, M. Rundgren, J. Bro-Jeppesen, J. Kjaergaard, C. Hassager, H. Lindehammar, J. Horn, S. Ullén, N. Nielsen, H. Friberg, and T. Cronberg. Time to epileptiform activity and eeg background recovery are independent predictors after cardiac arrest. *Clinical Neurophysiology*, 129:1660–1668, 2018.

[20] Jerry P. Nolan, Claudio Sandroni, Bernd W. Böttiger, Alain Cariou, Tobias Cronberg, Hans Friberg, Cornelia Genbrugge, Kirstie Haywood, Gisela Lilja, Véronique R.M. Moulaert,

Nikolaos Nikolaou, Theresa Mariero Olasveengen, Markus B. Skrifvars, Fabio Taccone, and Jasmeet Soar. European resuscitation council and european society of intensive care medicine guidelines 2021: Post-resuscitation care. *Resuscitation*, 161:220–269, 2021.

[21] Giulio Disanto, Michele Villa, Aleksandra Maleska Maceski, Chiara Prosperetti, Claudio Gobbi, Jens Kuhle, Tiziano Cassina, and Pamela Agazzi. Longitudinal serum neurofilament light kinetics in post- anoxic encephalopathy. *Annals of Clinical and Translational Neurology*, 10:2407–2412, 2023.

[22] Martin A. S. Meyer, Rasmus P. Beske, Simon Mølstrøm, Johannes Grand, Laust E. R. Obling, Sebastian Wiberg, Britt Borregaard, Simon Schneekloth, Sif Grau Kaad, Pernille M. Christensen, Christina Christoffersen, Ruth Frikke-Schmidt, Henrik Schmidt, Jacob E. Møller, Jesper Kjaergaard, and Christian Hassager. Neurofilament light chain for prognostication after cardiac arrest-first steps towards validation. *Critical Care*, 29:348, 8 2025.

[23] Wen Wu, Amit Chopra, Carolyn Ziegler, Shelley L. McLeod, and Steve Lin. 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.

[24] M.C. Tjepkema-Cloostermans, C. da Silva Lourenço, B.J. Ruijter, S.C. Tromp, G. Drost, F.H.M. Kornips, A. Beishuizen, F.H. Bosch, J. Hofmeijer, and M.J.A.M. van Putten. Outcome prediction in postanoxic coma with deep learning. *Critical care medicine*, 47, 2019.

[25] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2623–2631. ACM, 2019.

[26] Lawrence D Brown, T Tony Cai, and Anirban Dasgupta. Interval estimation for a binomial proportion. *Statistical Science*, 16:101–133, 2001.

[27] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *Proceedings of the 34 th International Conference on Machine Learning*. PMLR 70, 2017.

Conflict of interest

MvP is co-founder of Clinical Science Systems, a manufacturer of EEG software (NeuroCenter EEG) Author contributions: Conceptualisation: MvP and MT. Coding: MvP and MT. Data analysis: MvP and MT. Writing first draft: MvP. Data collection: all authors. Critical reading and commenting on the first draft: all authors.

A: Supplementary Methods

A.1. EEG preprocessing

Raw EEGs were recorded at 500 or 512 Hz using the international 10–20 system. Signals were filtered with a 1–25 Hz FIR bandpass filter and downsampled to 64 Hz. Data were detrended and segmented into non-overlapping 10-second epochs. 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 CRI value. If insufficient clean data were available, no CRI was computed for that time window.

A.2. Labeling and exclusion criteria

Outcome labels for the training data were based on CPC scores at 3 or 6 months post-discharge. Outcome labels for the validation set were based on patient records and CPC at hospital discharge was used. This has been shown to correlate well with long-term outcome [23]. Patients without follow-up or with non-neurological causes of death (e.g., hemodynamic failure or reinfarction) were excluded.

A.3. Model architecture

The CNN consisted of 3 convolutional layers with ReLU activations, batch normalization, dropout (0.3–0.5), and 1–2 dense layers (64–128 units). Hyperparameters included kernel sizes (3–11), L2 regularization (10^{-5} to 10^{-3}), and learning rates (10^{-4} to 10^{-2}). The Adam optimizer was used with early stopping and learning rate reduction on plateau.

A.4. Hyperparameter optimization

We used Optuna with 100 trials. For each trial, the model was trained on four of five stratified cross-validation folds. Training used a custom loss function combining binary cross-entropy with penalties for miscalibration, low specificity, and insufficient separation of the predicted probabilities. The relative weights of these penalties and the decision margin were included as tunable hyperparameters. Trial performance was evaluated using a composite Optuna objective (ROC AUC - calibration weight $\times$ ECE), encouraging the selection of models that were both discriminative and well calibrated.

A.5. Lock-in mechanism

Dynamic thresholds for good and poor outcome zones were derived from model trajectories on the training set and modeled as exponentials. For the boundary of the good region, we found

$$\mathrm{Thr}_{\mathrm{good}}(t) = 1.13\text{–}0.52\exp(-t/37.8) \tag{.1}$$

and for the poor

$$\mathrm{Thr}_{\mathrm{poor}}(t) = 0.08\text{–}0.10\exp(-t/5.7) \tag{.2}$$

where t is the time in hours. Values were clipped to [0, 1]. A lock was triggered if at least 8 CRI samples within a 60-minute window fell in the same zone; samples outside that zone (gray) were ignored. Poor-outcome locks were restricted to ≥ 12 h post-arrest.

A.6. Confidence intervals and calibration

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 0 or 1 [26]. Calibration was assessed using the expected calibration error (ECE) as described in [27]. Calibration metrics were not applied to the lock-in mechanism, as dynamic locking produces categorical rather than probabilistic outputs, precluding standard calibration analysis.

Table .2: Performance of DeepCRI in the external validation cohort without artefact rejection. Outcomes were locked after ≥ 8 non-gray zone predictions within 60 min; poor-outcome locks were restricted to ≥ 12 h post-arrest. Traditional ROC analysis at 24 h is included for comparison.

Metric	**Value**	**95% CI**
Total patients	223	–
Patients classified (locked)	199 (89%)	–
Gray zone patients	24 (11%)	–
For good outcome		
Sensitivity	96.4%	91.8–98.4%
Specificity	77.0%	67.1–84.6%
For poor outcome		
Sensitivity	62.6%	53.1–71.2%
Specificity	99.2%	95.9–99.8%
False predictions of good outcome	20	–
False predictions of poor outcome	1	–
ROC AUC at 24 h (non-locking)	0.89	0.86–0.93

B: Performance metrics without artefact rejection

In table .2 we show the performance metrics for the DeepCRI without artifact rejection.
The total number of EEGs evaluated is one more than reported in Table 1, as one EEG was excluded from that analysis as a result of too many artifacts.

C: Supplementary videos

Demonstration of the DeepCRI integrated into the NeuroCenter EEG platform, showing real-time prognostic output during continuous EEG monitoring in comatose post–cardiac arrest patients.