University of Groningen

Outcome Prediction in Postanoxic Coma With Deep Learning

Tjepkema-Cloostermans, Marleen C; da Silva Lourenço, Catarina; Ruijter, Barry J; Tromp, Selma C; Drost, Gea; Kornips, Francois H M; Beishuizen, Albertus; Bosch, Frank H; Hofmeijer, Jeannette; van Putten, Michel J A M

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Link to publication in University of Groningen/UMCG research database

Outcome Prediction in Postanoxic Coma With Deep Learning

Marleen C. Tjepkema-Cloostermans, PhD; Catarina da Silva Lourenço, BSc[2,3];

Barry J. Ruijter, MD, PhD[2]; Selma C. Tromp, MD, PhD[4]; Gea Drost, MD, PhD[5];

Francois H. M. Kornips, MD[6]; Albertus Beishuizen, MD, PhD[7]; Frank H. Bosch, MD, PhD[8];

Jeannette Hofmeijer, MD, PhD[2,9]; Michel J. A. M. van Putten, MD, PhD[1,2]

Objectives: Visual assessment of the electroencephalogram by experienced clinical neurophysiologists allows reliable outcome prediction of approximately half of all comatose patients after cardiac arrest. Deep neural networks hold promise to achieve similar or even better performance, being more objective and consistent.
Design: Prospective cohort study.
Setting: Medical ICU of five teaching hospitals in the Netherlands.
Patients: Eight-hundred ninety-five consecutive comatose patients after cardiac arrest.
Interventions: None.

Measurements and Main Results: Continuous electroencephalogram was recorded during the first 3 days after cardiac arrest. Functional outcome at 6 months was classified as good (Cerebral Performance Category 1−2) or poor (Cerebral Performance Category 3−5). We trained a convolutional neural network, with a VGG architecture (introduced by the Oxford Visual Geometry Group), to predict neurologic outcome at 12 and 24 hours after cardiac arrest using electroencephalogram epochs and outcome labels as inputs. Output of the network was the probability of good outcome. Data from two hospitals were used for training and internal validation ($n = 661$). Eighty percent of these data was used for training and cross-validation, the remaining 20% for independent internal validation. Data from the other three hospitals were used for external validation ($n = 234$). Prediction of poor outcome was most accurate at 12 hours, with a sensitivity in the external validation set of 58% (95% CI, 51−65%) at false positive rate of 0% (CI, 0−7%). Good outcome could be predicted at 12 hours with a sensitivity of 48% (CI, 45−51%) at a false positive rate of 5% (CI, 0−15%) in the external validation set.
Conclusions: Deep learning of electroencephalogram signals outperforms any previously reported outcome predictor of coma after cardiac arrest, including visual electroencephalogram assessment by trained electroencephalogram experts. Our approach offers the potential for objective and real time, bedside insight in the neurologic prognosis of comatose patients after cardiac arrest. (*Crit Care Med* 2019; XX:00−00)
Key Words: brain hypoxia; cardiac arrest; deep neural networks; electroencephalography; machine learning

More than half of the patients who remain comatose after cardiac arrest never regain consciousness (1). Early and accurate prediction of neurologic outcome supports clinical decision-making and may limit futile treatments of patients without relevant recovery perspectives (1–3). Furthermore, early knowledge about potential neurologic recovery is of high importance for family members (4). However, in current standard care, reliable prediction of poor

outcome is possible in only 20% of patients, based on neurologic examination and absent N20 responses in median nerve somatosensory evoked potential (SSEP) testing (4).

Visual assessment of early electroencephalogram (EEG) recordings has shown to comprise substantial additional value allowing reliable prediction of poor neurologic outcome in approximately half of all comatose patients after cardiac arrest (1, 5–8). Further, EEG allows for prediction of good neurologic outcome. In particular, suppressed EEG patterns at 24 hours after cardiac arrest or later or burst-suppression patterns with identical or synchronous bursts are invariably associated with poor outcome (1, 5–7). Otherwise, return of a continuous EEG pattern within 12 hours after cardiac arrest is a strong predictor of good outcome (1, 5–7). In addition to the analysis of the EEG background activity, modulation of EEG patterns in response to an external stimulus, EEG reactivity, has been explored (8–10). However, the potential relevance of reactivity in addition to assessment of the EEG background pattern has not been studied (11) and the reproducibility appears limited (12).

An important drawback of visual, qualitative analysis is the inability to capture the integral richness of the EEG signal. Furthermore, visual analysis by an experienced electroencephalographer is time consuming, requires a long training period (13) and suffers from both intra- and interobserver variability (14). Computer-assisted interpretation of EEG may overcome these limitations, allowing continuous assessment of the EEG, with promise of better discrimination between patients with various outcomes, higher consistency, and lower costs.

We previously reported on the Cerebral Recovery Index (CRI) to support visual assessment of the EEG for outcome prediction of patients with a postanoxic coma, using five quantitative EEG features, including continuity, amplitude, and frequency content (15). Significant improvement of the CRI was realized by adding more quantitative EEG features and using a random forest classifier (16). The CRI allows prediction of poor or good outcome as reliable as visual EEG assessment and SSEP, at higher sensitivity.

In this study, we take a different approach by using a deep convolutional neural network, avoiding the need for explicit feature extraction (16, 17) and using the unique ability from deep neural networks to learn from data (18–20). Deep learning has proven its potential for general and highly variable tasks, such as speech recognition and image grouping (19, 21). Recently, classification of skin lesions with a convolutional neural network trained on a large dataset of clinical images achieved performance on par with experienced dermatologists (20). Motivated by these successful approaches, we train a neural network end-to-end directly from the EEG and the corresponding labels, to predict neurologic outcome of comatose patients after cardiac arrest.

MATERIALS AND METHODS

In this prospective cohort study, all consecutive adult comatose (Glasgow Coma Scale score ≤ 8) patients after cardiac arrest admitted at the ICUs of five large teaching hospitals in the Netherlands (Medisch Spectrum Twente, Rijnstate, St. Antonius Hospital, University Medical Center Groningen, and VieCuri Medical Center) were included. Exclusion criteria were other types of severe brain injury (stroke, traumatic brain injury, or progressive neurodegenerative disease), either preexisting or coinciding with cardiac arrest. As EEG is part of routine care in all five centers, the Medical Ethical Committee Twente waived the need for informed consent for continuous EEG monitoring. Oral informed consent was obtained from surviving patients at the time of follow-up at 3 and 6 months.

Patients were treated according to standard protocols for comatose patients after cardiac arrest. Targeted temperature management (33°C or 36°C) was induced as soon as possible and maintained for 24 hours. Patients received propofol, midazolam, or both for sedation and morphine, fentanyl, or remifentanil for analgesia. In one center, the majority of patients was anesthetized with sevoflurane instead of propofol or midazolam. Withdrawal of treatment was considered during normothermia, off sedation, and later than 72 hours after cardiac arrest. Decisions on treatment withdrawal were based on international guidelines including bilateral absence of the SSEP, absent or extensor motor responses, and absence of brainstem reflexes (2, 22). Decisions on treatment withdrawal were sporadically taken between 48 and 72 hours in case of absent brainstem reflexes or SSEP responses. The EEG recorded in the first 72 hours after cardiac arrest was not taken into account in decision-making. More details on the treatment protocol and main outcomes regarding the predictive value of EEG (based on visual assessment) of the first 388 patients who were included in this study have been previously published (23). Data of the first 283 patients of two hospitals (Medisch Spectrum Twente and Rijnstate hospital) were used for the development of the CRI (16).

Outcome Assessment

Primary outcome measure was neurologic outcome at 6 months after cardiac arrest defined as the score on the Cerebral Performance Category (CPC) scale, dichotomized as good (CPC 1–2, no or mild neurologic impairment) or poor (CPC 3–5, severe neurologic impairment, vegetative state, or death). Outcome was assessed by a standardized telephone interview by one of two investigators (M.C.T.-C. or B.J.R.) or a trained research nurse.

Continuous EEG Recordings

Continuous EEG recordings were started as soon as possible after ICU admission, typically within 12–24 hours after cardiac arrest, and continued up to 3–5 days, unless patients regained consciousness or died at an earlier stage. EEGs were recorded with 21 silver/silver chloride cup electrodes placed on the scalp according to the international 10–20 system. A computer algorithm, as used in a previous quantitative EEG study (15), was used to select 5-minute artifact-free epochs at 12 and 24 hours after cardiac arrest. If no epoch was available at these time points, because of artifacts, the closest available artifact-free epoch in the range ± 2 hours was used.

Visual Analysis

Visual analysis of EEG data was performed offline. Before visual assessment, EEG data were band-pass filtered in the 0.5–35 Hz frequency range and re-referenced to the longitudinal bipolar montage. EEG epochs were presented in random order to reviewers who were blinded to the timing of the epoch, the clinical condition of the patients, medication, and outcome. All EEG epochs were assessed by two experienced reviewers from a pool of six (M.C.T.-C., B.J.R., M.J.A.M.v.P., H.K., A.G., or J.H.), independently. If two reviewers disagreed, the final classification was determined by consensus. If necessary, a third reviewer was consulted. Reviewers were allowed to choose the option "No classification possible" if the epoch was considered unreliable due to artifacts.

EEG patterns were classified as generalized suppression (all activity < 10 μV), low voltage (activity between 10 and 20 μV), burst suppression (≥ 50% suppression), discontinuous (10–50% suppression), continuous (< 10% suppression), or epileptiform. Burst suppression was further specified as synchronous burst-suppression patterns (burst suppression with generalized, sharp-onset bursts, with suppressed background; this includes burst suppression with identical bursts) or other burst-suppression patterns (including spatially heterogeneous burst-suppression patterns). Continuous patterns were subdivided according to their dominant frequency (< 4, 4–8, or > 8 Hz). Epileptiform patterns were further subdivided as generalized periodic discharges (GPDs) on a suppressed background (background activity < 10 μV), GPDs on a nonsuppressed background or other epileptiform patterns (including electrographic seizures).

Convolutional Neural Network

Training and validation of the neural network was performed using the 5-minute epochs at 12 and 24 hours after cardiac arrest. Epochs were partitioned into 10-second nonoverlapping fragments, resulting in 30 fragments at a given time point for each patient. Before presenting to the network, data were band-pass filtered in the 0.3–25 Hz frequency range, downsampled to 64 Hz and re-referenced to 1) the longitudinal bipolar montage (using 19 electrodes) and 2) the Laplacian montage (using 19 electrodes).

Data from two hospitals (Medisch Spectrum Twente and Rijnstate hospital) were used for training, cross-validation, and internal validation. Data of 80% of these patients were randomly selected and used for training, the remainder for internal validation. Training was performed with 10-fold cross-validation. Data of the remaining three centers (St. Antonius Hospital, University Medical Center Groningen, and VieCuri Medical Center) were used for external validation to assess the general applicability of the models.

A deep learning convolutional neural network was implemented in Python using Keras and Theano and a computer unified device architecture-enabled NVIDIA GPU (GTX-1080) (NVIDIA, Santa Clara, CA), running on CentOS 7. The architecture of the network is shown in **Supplemental Figure 1** (Supplemental Digital Content 1, http://links.lww.com/CCM/E643

TABLE 1. Patient Characteristics and Medication Use in Patients With Good and Poor Outcomes

Characteristics	Good Outcome (n = 397)	Poor Outcome (n = 467)
Sex, male, n (%)	312 (79)	343 (73)
Age, yr, mean ± SD	60±13	65±14
Location, n (%)		
Out-of-hospital cardiac arrest	369 (93)	419 (90)
In-hospital cardiac arrest	28 (7)	48 (10)
Initial electrocardiogram rhythm, n (%)		
VF/VT rhythm	360 (91)	264 (63)
Non VF/VT rhythm	30 (8)	176 (38)
Unknown	7 (2)	27 (6)
Presumed cause of cardiac arrest, n (%)		
Cardiac etiology	353 (89)	321 (69)
Noncardiac etiology	21 (5)	97 (21)
Unknown	23 (6)	49 (10)
Targeted temperature management, n (%)	391 (98)	440 (94)
Hypothermia (32–34°C), n (%)	179 (45)	220 (47)
Treated with propofol, n (%)	354 (89)	388 (83)
Dose (mg/kg/hr), mean ± SD	3.2±1.2	2.8±1.1
Treated with midazolam, n (%)	105 (26)	135 (29)
Dose (μg/kg/hr), mean ± SD	118±71	122±88
Treated with fentanyl, n (%)	156 (39)	216 (46)
Dose (μg/kg/hr), mean ± SD	1.7±0.8	1.4±0.8
Treated with remifentanil, n (%)	21 (5)	33 (7)
Dose (μg/kg/hr), mean ± SD	7.2±4.5	4.4±3.1
Treated with morphine, n (%)	193 (46)	174 (37)
Dose (μg/kg/hr), mean ± SD	27±11	29±14
Treated with sevoflurane, n (%)	21 (5)	30 (6)
End-tidal volume %, mean ± SD	1.4±0.3	1.3±0.3
Somatosensory evoked potential performed, n (%)	43 (11)	280 (60)
N20 bilaterally absent, n (%)	0 (0)	124 (27)

VF = ventricular fibrillation, VT = ventricular tachycardia.

and based on a VGG model C network, created in 2014 in Oxford by Simonyan and Zisserman (24) from the Visual Geometry Group. This architecture was chosen because of its success in detection and image classification (25), including health-related problems (26, 27). Stochastic optimization

was realized using Adam (learning rate = 0.00002, β_1 = 0.91, β_2 = 0.999, and ϵ = 10^{-8}) (28). As the loss function, the binary cross-entropy was used. Outcome of the network was the probability of good neurologic outcome, defined as the mean probability of the 30 EEG fragments in the 5-minute recording.

We trained separate networks for the two different montages at 12 and 24 hours after cardiac arrest. We also studied the prognostic value by using information from the temporal evolution of the EEG, combining data from 12 and 24 hours after cardiac arrest. To this end, we trained the neural network with 18-channel 10-second EEG fragments, containing both recordings obtained at 12 hours (n = 9 channels) and at 24 hours (n = 9 channels) after cardiac arrest.

Cerebral Recovery Index

To compare the results of the deep learning model with our recently published CRI (16), we calculated this index for all patients in the external validation set group as well.

Statistical Analyses and Creation of Receiver Operating Characteristic Curves

After training, classification performance was evaluated in the independent internal validation set as well as in the external validation set using Matlab R2017a (The MathWorks, Natick,

MA). Predictions for poor and good outcomes are represented as receiver operating characteristic (ROC) curves. Based on these ROC curves, the montage with the highest accuracy (largest area under the curve [AUC]) was chosen, for which two thresholds were set, one for predicting poor neurologic outcome and the other for predicting good neurologic outcome. For the prediction of poor outcome, we only considered threshold values at zero false positive rate (FPR). For the prediction of good outcome, we allowed for a maximum of 5% false positives. The sensitivity, FPR, and positive and negative predictive value set at these thresholds were calculated for the validation set with their corresponding 95% CIs.

RESULTS

Of the 895 patients who were included, 31 were lost to follow-up. Of the remaining patients, 397 (46%) had good neurologic outcome. Patient characteristics are presented in **Table 1**.

Training and Internal Validation

The internal set consisted of 661 patients after cardiac arrest from the Medisch Spectrum Twente and Rijnstate hospital. At 12 hours after cardiac arrest, EEG epochs of 374 patients were available, divided over a training (n = 300) and internal validation set (n = 74). At 24 hours after cardiac arrest, EEG epochs
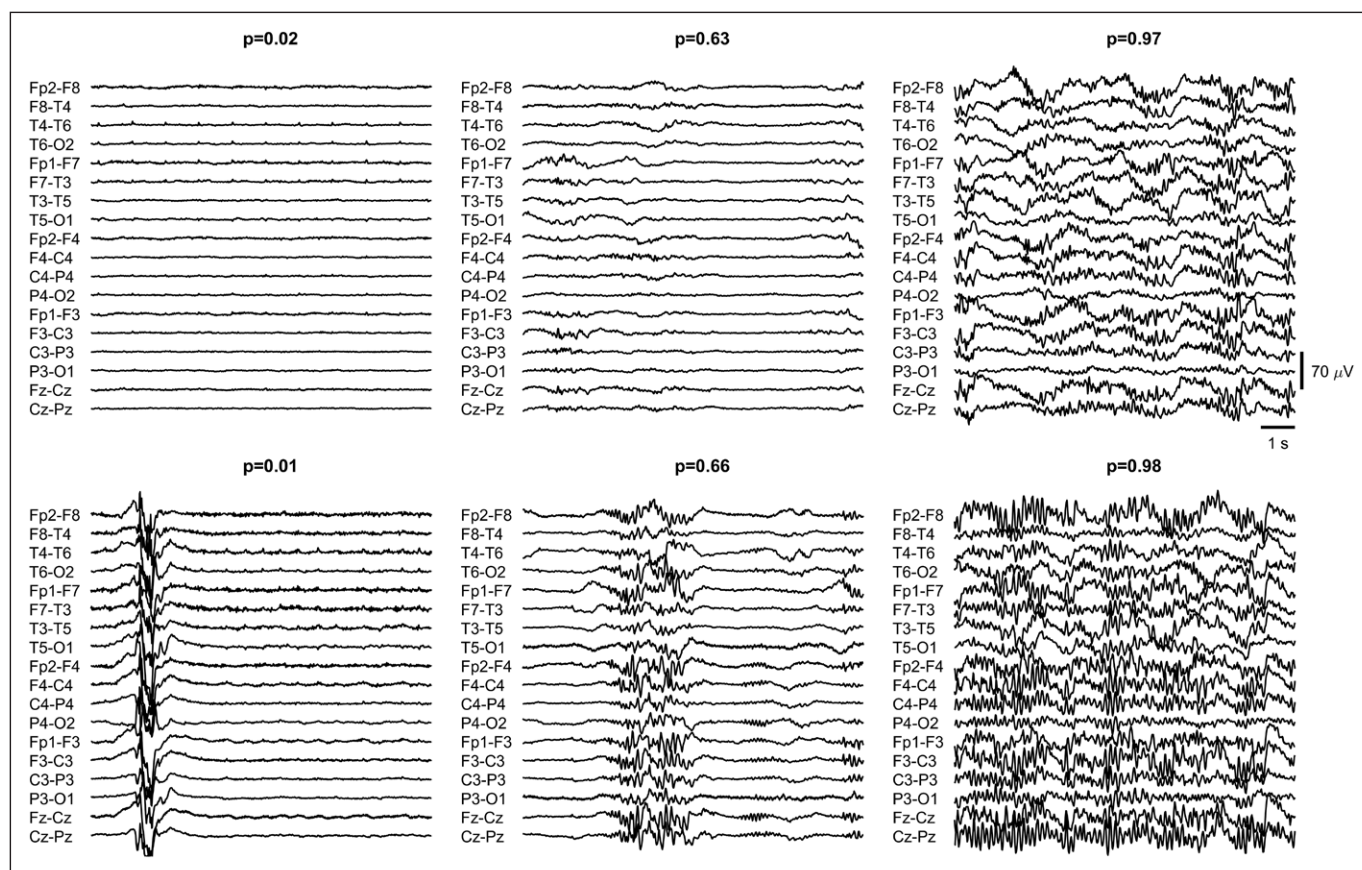
Figure 1. Examples of electroencephalogram (EEG) fragments of six different patients at 12 hr after cardiac arrest. The probability of good outcome predicted by the deep neural network is shown above each panel. *Left*: Two EEG fragments, classified as "unfavorable," both patients indeed had a poor neurologic outcome and died (Cerebral Performance Category [CPC] = 5). *Center*: Two EEG fragments, classified as "uncertain" neurologic outcome. *Right*: Two EEG fragments that are classified as "favorable," both patients had good neurologic outcome (CPC = 1).

of 534 patients were available, again divided over a training ($n = 428$) and validation set ($n = 106$). More EEG epochs were available at 24 hours than at 12 hours after cardiac arrest since continuous EEG recordings were often started within 12–24 hours after cardiac arrest. In 342 patients, EEG epochs at both time points (12 and 24 hr) after cardiac arrest were available. Combinations of these epochs were also distributed over a training ($n = 274$) and a validation set ($n = 68$).

Training of the convolutional neural network with up to 7772 EEG samples of 10 seconds took approximately 80 minutes. Examples of 10-second EEG fragments at 12 hours after cardiac arrest are shown in **Figure 1**, including the probabilities for good outcome predicted by the deep neural network.

The ROC curves for the prediction of poor outcome in the internal test and internal validation set are shown in **Figure 2**. Similar accuracies were obtained for 12 and 24 hours after cardiac arrest, with AUCs of 0.86–0.90. Combining data from 12 and 24 hours did not increase the predictive accuracy. Comparison between the longitudinal bipolar and Laplacian showed no statistically significant differences in the AUC. Therefore, for further analyses, the longitudinal bipolar montage was used. In the internal independent validation set, sensitivities for the prediction of poor outcome were 42% and 57% at 12 and 24 hours after cardiac arrest, respectively, both at a FPR of 0%. Good outcome could be predicted with sensitivities
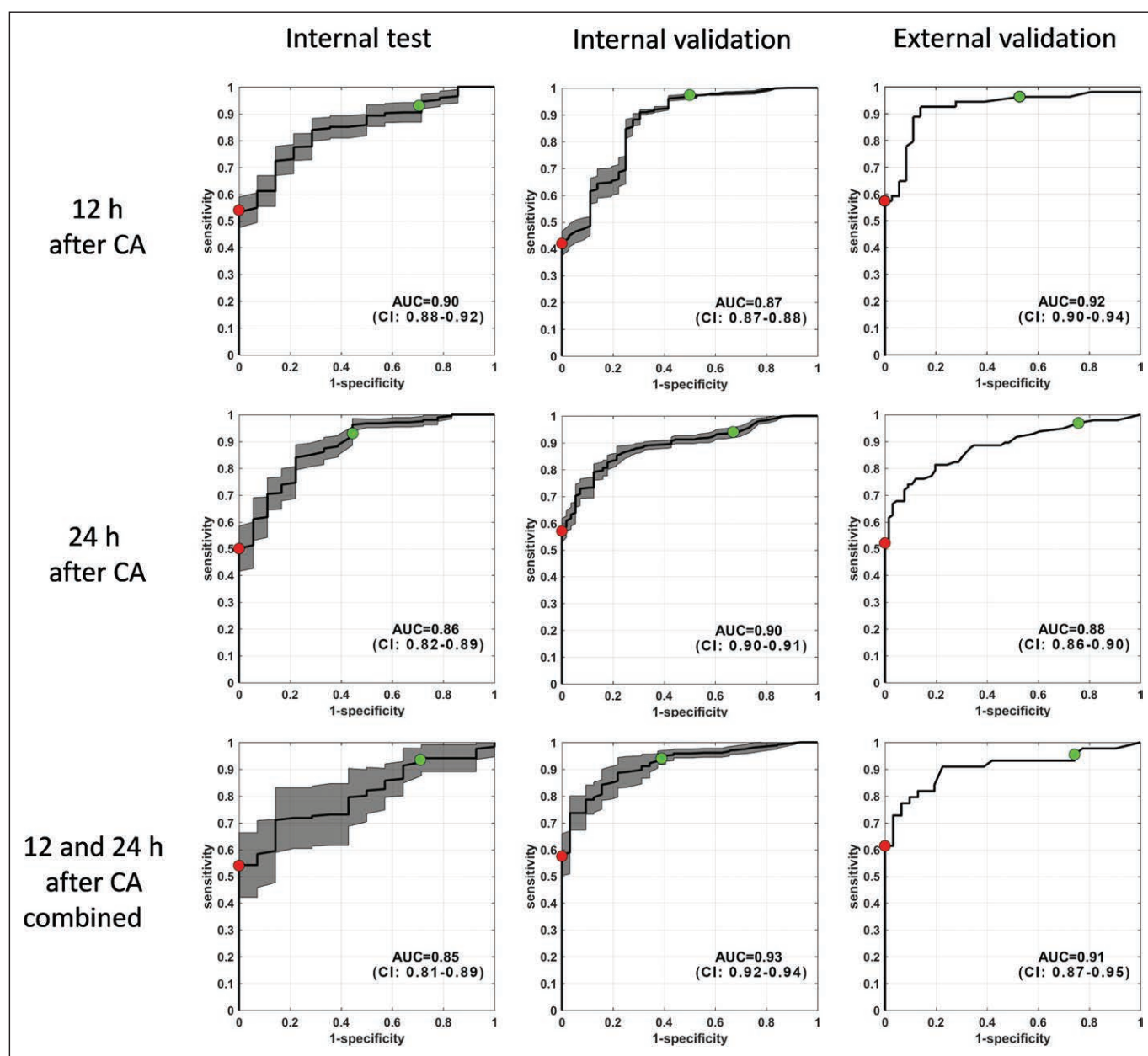
Figure 2. Receiver operating characteristic (ROC) curves for the prediction of poor outcome (Cerebral Performance Category [CPC] 3–5) in the internal test set (*left*), internal validation set (*center*), and external validation set (*right*). Shown are ROC curves at 12 hr after cardiac arrest (CA) (*top*), at 24 hr after CA (*middle*), and at 12 and 24 hr combined (*bottom*). The **solid red** and **green circles** indicate the chosen threshold for the prediction of poor and good outcomes, respectively. The *gray area* indicates the 95% CI. AUC = area under the curve.

of 48% and 33% at 12 and 24 hours after cardiac arrest, respectively, both at a FPR of 5% (**Table 2**).

External Validation

The external validation set consisted of 234 patients after cardiac arrest from the St. Antonius Hospital, University Medical Center Groningen, and VieCuri Medical Center. At 12 hours after cardiac arrest, EEG epochs of 91 patients were available. At 24 hours after cardiac arrest, EEG epochs of 163 patients were available. In 76 patients, EEG epochs at both time points (12 and 24 hr) after cardiac arrest were available.

The ROC curves for the prediction of poor outcome in the external validation set are shown in Figure 2-right column. In this external validation set, outcome prediction was most accurate at 12 hours after cardiac arrest. For the prediction of poor outcome, sensitivity was 58% at a FPR of 0% at 12 hours after cardiac arrest. Good outcome could be predicted with a sensitivity of 48% at a FPR of 5% at 12 hours after cardiac arrest. Whereas at 24 hours after cardiac arrest, sensitivity for poor outcome was 51% at a FPR of

0%, the sensitivity for good outcome was 22% at a FPR of 5% (Table 2).

The comparison of the performance of the deep learning algorithm with the CRI in the external validation set is shown in **Supplemental Figure 2** (Supplemental Digital Content 2, http://links.lww.com/CCM/E644). At 12 hours after cardiac arrest, the performance of the deep learning algorithm and CRI are almost equal (both AUC of 0.92), whereas at 24 hours, the deep learning algorithm outperforms the CRI (AUC of 0.88 vs 0.79).

Visual Assessment

In the complete dataset (internal test, internal validation, and external validation set combined), EEG patterns that were classified as suppressed, synchronous burst suppression, or GPDs on a suppressed background were invariably associated with a poor neurologic outcome, with a sensitivity of 37% (CI, 31–44%) and 25% (CI, 20–29%), respectively, at 12 and 24 hours after cardiac arrest, both with a FPR of 0% (CI, 0–2% and 0–1%). A continuous EEG pattern (irrespective of frequency content) was associated with good neurologic outcome with sensitivity

TABLE 2. Predictive Values Including Corresponding 95% CIs for the Prediction of Both Poor and Good Outcomes in the Internal Test Set, Internal Validation Set, and External Validation Set Using the Longitudinal Bipolar Montage

Time After Cardiac Arrest	Dataset	Sensitivity (CI)	False Positive Rate (CI)	Positive Predictive Value (CI)	Negative Predictive Value (CI)
Prediction of poor outcome					
12 hr	Internal test	56% (44–68%)	0% (0–12%)	100% (95–100%)	70% (64–76%)
	Internal validation	42% (36–48%)	0% (0–8%)	100% (99–100%)	62% (60–65%)
	External validation	58% (51–65%)	0% (0–7%)	100% (96–100%)	61% (59–63%)
24 hr	Internal test	50% (41–59%)	0% (0–12%)	100% (95–100%)	65% (60–71%)
	Internal validation	57% (54–60%)	0% (0–10%)	100% (96–100%)	69% (64–74%)
	External validation	51% (49–53%)	0% (0–5%)	100% (97–100%)	59% (58–60%)
12 and 24 hr combined	Internal test	54% (41–67%)	0% (0–11%)	100% (94–100%)	68% (63–73%)
	Internal validation	58% (51–65%)	0% (0–9%)	100% (96–100%)	67% (64–70%)
	External validation	61% (59–63%)	0% (0–8%)	100% (92–100%)	64% (61–67%)
Prediction of good outcome					
12 hr	Internal test	30% (22–38%)	5% (0–15%)	90% (87–93%)	68% (64–73%)
	Internal validation	48% (45–51%)	5% (0–13%)	94% (93–95%)	67% (65–68%)
	External validation	48% (45–51%)	5% (0–15%)	89% (85–93%)	73% (70–76%)
24 hr	Internal test	56% (50–61%)	5% (0–18%)	90% (87–94%)	67% (62–72%)
	Internal validation	33% (30–36%)	5% (0–19%)	81% (75–87%)	67% (65–69%)
	External validation	22% (20–25%)	5% (0–14%)	65% (61–69%)	84% (79–90%)
12 and 24 hr combined	Internal test	29% (20–39%)	5% (0–17%)	79% (75–82%)	64% (62–66%)
	Internal validation	61% (60–62%)	5% (0–13%)	92% (91–93%)	78% (76–79%)
	External validation	27% (24–31%)	5% (0–15%)	65% (61–69%)	80% (76–84%)
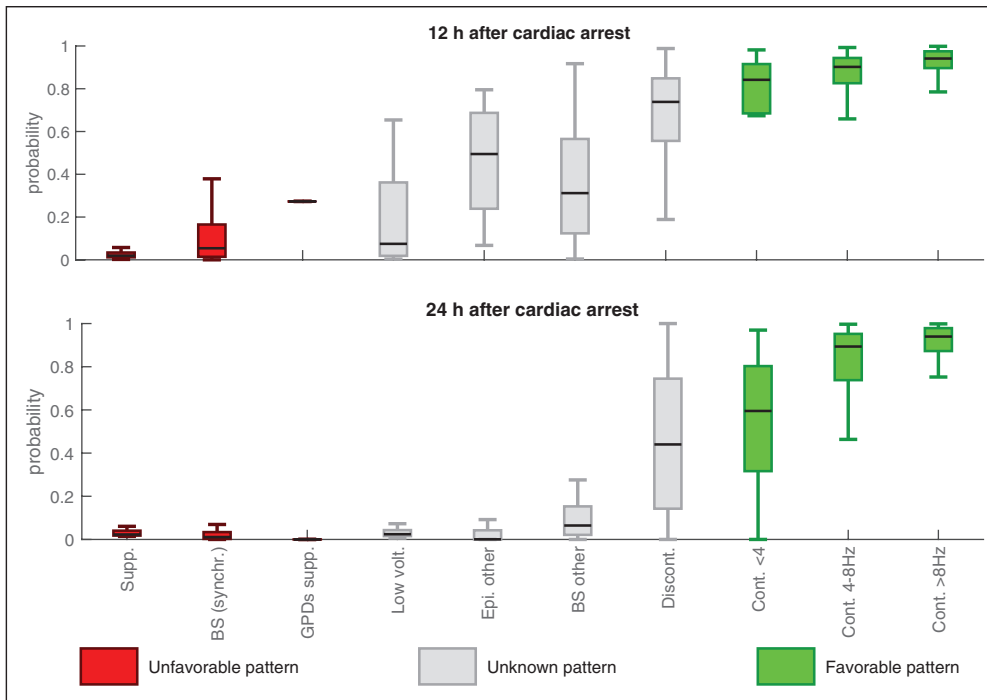
Figure 3. *Box plots* showing the deep neural network assigned probability for a good outcome, grouped according to the electroencephalogram (EEG) pattern determined by visual interpretation. On each *box plot*, the central mark (*thick black line*) indicates the median, and the bottom and top edges of the box indicate the 25–75th percentiles, respectively. The *whiskers* extend to the most extreme data points not considered outliers. Note that the EEG patterns that are visually classified as "favorable" (in *green*) indeed have much higher probabilities of good outcome than "unfavorable" EEG patterns (in *red*). BS (synchr.) = burst suppression with bilateral synchrony and suppressed interburst intervals, BS other = other burst-suppression patterns, including spatially heterogeneous bursts with gradual transitions, Cont. = continuous, Discont. = discontinuous, Epi. other = epileptiform, other than GPDs on a suppressed background, GPDs supp. = generalized periodic discharges on a suppressed background, Low volt. = low voltage (10–20 μV), Supp. = suppressed (< 10 μV).

CRI (16), which was based on combinations of handmade quantitative EEG features. In all other patients, instead of a classification as "good," "intermediate," or "poor," the deep learning algorithm presents a quantitative probability of good outcome. In patients with intermediate recovery perspectives, this can contribute to a multimodal decision process (3).

To the best of our knowledge, this is the first large-scale application of deep learning on EEG data from comatose patients after cardiac arrest. With this approach, the algorithm "learns" particular characteristics of the data that are difficult or even impossible to assess by a human expert and overcomes limitations associated with defining "handmade" quantitative features. Other benefits include extremely fast processing, allowing real-time bedside application, and use by non-EEG experts such as intensivists. As human expertise for trustworthy visual assessment of these EEG recordings is not available in all medical centers, deep learning can extend this to these clinics and allows 24/7 reliable and consistent interpretation of the EEG. An important strength of the current study is that we validated our network on an external validation set with EEG data from three medical centers.

A limitation is that feature discovery is not straightforward, as the convolutional neural network does not directly explicate which particular EEG features are recognized as favorable or unfavorable (30). Although various techniques exist to probe deep neural networks to elucidate the features used, thus allowing feature discovery (31), this is outside the scope of the current work. EEG interpretation in this study may have been influenced by the use of sedative medication. However, previous studies demonstrated that visual analysis of the EEG is reliable in these patients, despite the use of mild therapeutic hypothermia and sedation. Specific unfavorable patterns (i.e., suppressed or burst suppression with identical bursts) cannot be solely induced by hypothermia, propofol, or midazolam in the relatively low doses that were used in these patients. Likewise, quantitative EEG measures seem to be more influenced by the anoxic encephalopathy itself than by the effects of sedation (32, 33). For primary outcome, we used the CPC score at 6 months, a few patients had full neurologic recovery but died due to a second cardiac arrest or another nonneurologic

of 50% (CI, 43–57%) and a FPR of 9% (CI, 6–14%) at 12 hours after cardiac arrest and a sensitivity of 68% (CI, 63–73%) at a FPR of 20% (CI, 16–24%) at 24 after cardiac arrest. The distribution of the probabilities given by the deep learning network for each visual EEG class is shown in **Figure 3**.

DISCUSSION

Visual assessment of the EEG background pattern by experienced clinical neurophysiologists provides reliable information for the prediction of poor or good neurologic outcome in up to half of all comatose patients after cardiac arrest (1, 5–7). Since qualitative, visual assessment may neglect a significant part of the information present in evolving EEG rhythms, we hypothesized that deep convolutional neural networks can extract substantially more information, in turn providing equally reliable predictions in more patients as compared with visual EEG. With this approach, we were indeed able to reliably predict neurologic outcome of comatose patients after cardiac arrest in substantially more patients than with visual EEG assessment. Predictive values were also higher in comparison to neurologic examination or SSEP (1, 6, 7, 29). We could predict poor outcome with a sensitivity of 58% at a FPR of 0% at 12 hours after cardiac arrest. Good neurologic outcome could be predicted at 12 hours after cardiac arrest with a sensitivity of 48% at a FPR of 5%. This performance equals that of our previously published

problem, which might have slightly increased the FPR for the prediction of good neurologic outcome.

Current performance might be further improved by using the full temporal evolution of the complete EEG recording, instead of analyzing brief 5-minute epochs (at 12 or 24 hr after cardiac arrest). Other network architectures, for example, networks with more convolutional layers or recurrent neural networks, may also result in an improvement in classification accuracy (34, 35).

Deep neural networks hold promise to assist in the interpretation of the EEG in other clinical conditions, ranging from sleep staging (36) to detection of epileptiform discharges (37) or seizures (38, 39), and it is foreseen that deep learning will significantly affect how physicians will assess EEG recordings in the near future.

CONCLUSIONS

We present a classifier for the prediction of neurologic outcome after cardiac arrest, providing fast, reliable, and objective prognostic information for all patients, with the potential to be used bedside. The classifier is based on the whole spectrum of information that is included in evolving EEG patterns extracted by a convolutional neural network. For almost 100 years, the gold standard in the clinic for the interpretation of the EEG has been visual analysis by human experts (40). We show that a trained convolutional neural network can perform this task with similar or even better prognostic accuracy in patients with a postanoxic coma.

REFERENCES

1. Hofmeijer J, Beernink TM, Bosch FH, et al: Early EEG contributes to multimodal outcome prediction of postanoxic coma. *Neurology* 2015; 85:137–143

2. Wijdicks EF, Hijdra A, Young GB, et al; Quality Standards Subcommittee of the American Academy of Neurology: Practice parameter: Prediction of outcome in comatose survivors after cardiopulmonary resuscitation (an evidence-based review): Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 2006; 67:203–210

3. Oddo M, Rossetti AO: Early multimodal outcome prediction after cardiac arrest in patients treated with hypothermia. *Crit Care Med* 2014; 42:1340–1347

4. Sandroni C, Cavallaro F, Callaway CW, et al: Predictors of poor neurological outcome in adult comatose survivors of cardiac arrest: A systematic review and meta-analysis. Part 2: Patients treated with therapeutic hypothermia. *Resuscitation* 2013; 84:1324–1338

5. Tjepkema-Cloostermans MC, Hofmeijer J, Trof RJ, et al: Electroencephalogram predicts outcome in patients with postanoxic coma during mild therapeutic hypothermia. *Crit Care Med* 2015; 43:159–167

6. Sivaraju A, Gilmore EJ, Wira CR, et al: Prognostication of post-cardiac arrest coma: Early clinical and electroencephalographic predictors of outcome. *Intensive Care Med* 2015; 41:1264–1272

7. Spalletti M, Carrai R, Scarpino M, et al: Single electroencephalographic patterns as specific and time-dependent indicators of good and poor outcome after cardiac arrest. *Clin Neurophysiol* 2016; 127:2610–2617

8. Rossetti AO, Rabinstein AA, Oddo M: Neurological prognostication of outcome in patients in coma after cardiac arrest. *Lancet Neurol* 2016; 15:597–609

9. Noirhomme Q, Lehembre R, Lugo Zdel R, et al: Automated analysis of background EEG and reactivity during therapeutic hypothermia in comatose patients after cardiac arrest. *Clin EEG Neurosci* 2014; 45:6–13

10. Tsetsou S, Novy J, Oddo M, et al: EEG reactivity to pain in comatose patients: Importance of the stimulus type. *Resuscitation* 2015; 97:34–37

11. Admiraal MM, van Rootselaar AF, Horn J: Electroencephalographic reactivity testing in unconscious patients: A systematic review of methods and definitions. *Eur J Neurol* 2017; 24:245–254

12. Hermans MC, Westover MB, van Putten MJAM, et al: Quantification of EEG reactivity in comatose patients. *Clin Neurophysiol* 2016; 127:571–580

13. Gaspard N: ACNS critical care EEG terminology: Value, limitations, and perspectives. *J Clin Neurophysiol* 2015; 32:452–455

14. Gerber PA, Chapman KE, Chung SS, et al: Interobserver agreement in the interpretation of EEG patterns in critically ill adults. *J Clin Neurophysiol* 2008; 25:241–249

15. Tjepkema-Cloostermans MC, van Meulen FB, Meinsma G, et al: A Cerebral Recovery Index (CRI) for early prognosis in patients after cardiac arrest. *Crit Care* 2013; 17:R252

16. Tjepkema-Cloostermans MC, Hofmeijer J, Beishuizen A, et al: Cerebral recovery index: Reliable help for prediction of neurologic outcome after cardiac arrest. *Crit Care Med* 2017; 45:e789–e797

17. Foreman B, Claassen J: Quantitative EEG for the detection of brain ischemia. *Crit Care* 2012; 16:216

18. Bengio Y, Courville A, Vincent P: Representation learning: A review and new perspectives. *IEEE Trans Pattern Anal Mach Intell* 2012; 35:1–34

19. LeCun Y, Bengio Y, Hinton G: Deep learning. *Nature* 2015; 521:436–444

20. Esteva A, Kuprel B, Novoa RA, et al: Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; 542:115–118

21. Angermueller C, Pärnamaa T, Parts L, et al: Deep learning for computational biology. *Mol Syst Biol* 2016; 12:878

22. Sandroni C, Cariou A, Cavallaro F, et al: Prognostication in comatose survivors of cardiac arrest: An advisory statement from the European Resuscitation Council and the European Society of Intensive Care Medicine. *Resuscitation* 2014; 85:1779–1789

23. Sondag L, Ruijter BJ, Tjepkema-Cloostermans MC, et al: Early EEG for outcome prediction of postanoxic coma: Prospective cohort study with cost-minimization analysis. *Crit Care* 2017; 21:111

24. Simonyan K, Zisserman A: Very deep convolutional networks for large-scale image recognition. arXiv:14091556v6. ICLR 2015, San Diego, CA, May 7-9, 2015

25. Shin HC, Roth HR, Gao M, et al: Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* 2016; 35:1285–1298

26. Bejnordi BE, Lin J, Glass B, et al: Deep learning-based assessment of tumor-associated stroma for diagnosing breast cancer in histopathology images. 2017 IEEE 14th International Symposium on Biomedical Imaging, ISBI 2017. Melbourne, VIC, Australia, IEEE Computer Society, 2017, pp 929–932

27. Ravi D, Wong C, Deligianni F, et al: Deep learning for health informatics. *IEEE J Biomed Health Inform* 2017; 21:4–21

28. Kingma DP, Lei Ba J: Adam: A method for stochastic optimization. 3rd International Conference for Learning Representations. San Diego, CA, 2015

29. Sandroni C, Cariou A, Cavallaro F, et al: Prognostication in comatose survivors of cardiac arrest: An advisory statement from the European Resuscitation Council and the European Society of Intensive Care Medicine. *Intensive Care Med* 2014; 40:1816–1831

30. Längkvist M, Karlsson L, Loutfi A: A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognit Lett* 2014; 42:11–24

31. Zeiler MD, Fergus R: Visualizing and understanding convolutional networks. *In*: Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, volume 8689. Fleet D, Pajdla T, Schiele B (Eds). Springer, Cham, 2013, pp 818–833

32. Ruijter BJ, Hofmeijer J, Tjepkema-Cloostermans MC, et al: The prognostic value of discontinuous EEG patterns in postanoxic coma. *Clin Neurophysiol* 2018; 129:1534–1543

33. Drohan CM, Cardi AI, Rittenberger JC, et al: Effect of sedation on quantitative electroencephalography after cardiac arrest. *Resuscitation* 2018; 124:132–137

34. Choi E, Schuetz A, Stewart WF, et al: Using recurrent neural network models for early detection of heart failure onset. *J Am Med Inform Assoc* 2017; 24:361–370

35. Schmidhuber J: Deep learning in neural networks: An overview. *Expert Syst Appl* 2014; 72:221–230

36. Chambon S, Galtier MN, Arnal PJ, et al: A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. *IEEE Trans Neural Syst Rehabil Eng* 2018; 26:758–769

37. Johansen AR, Jin J, Maszczyk T, et al: Epileptiform spike detection via convolutional neural networks. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Shanghai, China, IEEE, 2016, pp 754–758

38. Thodoroff P, Pineau J, Lim A: Learning robust features using deep learning for automatic seizure detection. 2016 Machine Learning for Healthcare Conference. Los Angeles, CA, 2016, pp 1–12

39. Wulsin DF, Gupta JR, Mani R, et al: Modeling electroencephalography waveforms with semi-supervised deep belief nets: Fast classification and anomaly measurement. *J Neural Eng* 2011; 8:036015

40. Berger H: Über das Elektrenkephalogramm des Menschen. *Arch Psychiatr Nervenkr* 1929; 87:527–570