

MSc Technical Medicine
Medical Sensing & Stimulation

Clinical Specialization Internship

**Early prognostication in postanoxic coma patients
based on EEG with reduced electrode sets**

Myrthe van Merkerk
s1467832

June 24, 2021

Supervisors:

M.J.A.M. van Putten, MD, PhD

A.F. van Rootselaar, MD, PhD

J. Horn, MD, PhD

M.M. Admiraal, PhD

E.M. Walter, MSc

Amsterdam UMC, location AMC & University of Twente

Preface

This thesis is the final product of a ten-month internship in the department of Clinical Neurophysiology in the Amsterdam UMC – location AMC. This internship was the final step towards graduation and obtaining my Master's degree in Technical Medicine at the University of Twente. During these months, I had the opportunity to develop myself both clinically and scientifically, but not the least also on a personal and professional level. I am very grateful for all the experiences that contributed to this development. Therefore, I would like to take a moment to express my gratitude to the people who helped and supported me during this period.

First of all, I would like to thank Marjolein Admiraal for keeping an eye on the progress of the project and checking in on me from time to time. As my daily supervisor, you were always available for questions and helped me keep the time schedule realistic. I would like to thank Fleur van Rootselaar, as my medical supervisor, for giving me the opportunity to develop my clinical skills and the confidence to let me grow as a healthcare professional. You always triggered me to keep thinking about the clinical relevance of the whole project. As my third supervisor in the AMC, I would also like to thank Janneke Horn for your input on practical aspects for application in the ICU and your enthusiasm and encouragement on the project. Not to forget, I would like to thank the three of you for creating time into your busy schedules to visually assess almost 250 EEG fragments.

As my technical supervisor from the University of Twente, I would like to thank Michel van Putten for sharing his knowledge and critical thinking. You pushed me to keep a critical eye on the execution of the analyses and the presentation of the results. Additionally, I would like to thank all the technicians of the Clinical Neurophysiology department for thinking along and always being willing to help. And not only practical, but you also made me feel welcome in the department and a part of the group from the very beginning of my internship. Furthermore, I would like to thank Erik-Jan Meulenbrugge for always being available to answer my questions about programming, even if I had asked them twice already. You lifted my analyses to a higher level. Elyse Walter, I would like to thank you for guiding my process during the past two years. You encouraged me to think about my personal and professional needs and made me work to clear this up for myself. Lastly, I would like to thank Erik Groot Jebbink for taking the time to read my thesis from an objective point of view as the external member of my graduation committee.

Finally, I would like to thank all my friends and family for their support along the way and for making my whole study time a lot of fun.

I hope you enjoy reading this thesis.

Myrthe van Merkerk

Utrecht - June 24, 2021

Abstract

INTRODUCTION: Continuous electroencephalography (cEEG) is increasingly used for prognostication after cardiac arrest (CA). The first 24 hours of EEG after resuscitation are reported to have the highest prognostic value for neurological outcome. Easy-to-apply adhesive electrodes, like the BrainStatus, or a reduced number of electrodes can both contribute to early start of cEEG monitoring. Furthermore, quantitative EEG analysis can be a helpful tool to support prediction of neurological outcome.

OBJECTIVE: During this study, we investigated the possibilities for using adhesive electrodes and reduced montages for cEEG monitoring while preserving the predictive value.

METHODS: Continuous EEG was obtained from postanoxic comatose patients after CA admitted to the ICU for all parts of the study. Monitoring with the BrainStatus was done prospectively, simultaneously with conventional EEG monitoring, and with intermittent check-ups. A visual scoring application was used to score EEG epochs from the BrainStatus. Secondly, a total of 221 EEG epochs was visually scored in a 4-channel frontotemporal montage. All scored classifications were compared to the classifications in the 9-channel bipolar montage. Lastly, a logistic regression (LR) model was used for the prediction of neurological outcome. Features extracted from reduced electrode montages were used as input, and model performance was compared to the full montage.

RESULTS: Classification agreement of EEG background patterns between the BrainStatus and the bipolar montage was moderate ($\kappa = 0.48$), based on six patients. Agreement between the frontotemporal montage and the bipolar montage was substantial ($\kappa = 0.76$). Sensitivity and specificity for predicting neurological outcome were not significantly different for the frontotemporal montage compared to the bipolar montage. Baseline performance of the LR model for the bipolar montage equals an AUC of 0.897 (0.885-0.908), Se100 of 77% (74-79) for predicting poor outcome and Se95 of 71% (68-74) for good outcome at 12 hours after CA. For 24 hours after CA, the AUC equals 0.879 (0.873-0.886), with Se100 of 63% (62-65) for predicting poor outcome and Se95 of 50% (47-52) for good outcome. Performance of the model on the frontotemporal montage significantly increases for Se100 at 12h after CA, and for AUC and Se100 at 24h after CA.

CONCLUSION: Intermittent check-ups did not improve the signal quality of the BrainStatus recordings. Reducing the number of electrodes from 9 to 4 does not affect EEG classification or prognostic accuracy in patients with postanoxic coma. For qEEG analysis, the 4-channel montage showed significant better performance for outcome prediction than the 9-channel montage.

Contents

Preface	2
Abstract	3
List of figures	6
List of tables	7
1 Introduction	8
2 Materials and methods	10
Study population	10
Outcome assessment	10
Standard of care and monitoring.....	10
Epoch selection and EEG montages	10
Visual scoring application.....	11
Inter-rater reliability	12
Classification agreement.....	12
2.1 <i>BrainStatus</i>	12
Study population and outcome assessment	12
Monitoring and check-ups.....	12
Visual scoring	12
Evaluation of signal quality	13
2.2 <i>Electrode reduction for visual assessment</i>	13
Study population and outcome assessment	13
Visual scoring	13
Outcome prediction	13
2.3 <i>Electrode reduction for qEEG analysis</i>	14
Study population and outcome assessment	14
Preprocessing and model input	14
Feature extraction	14
Logistic regression model	14
Performance metrics	14
Predetermined reduced sets	15
L1 regularization.....	15
3 Results	16
3.1 <i>BrainStatus</i>	16

Patient characteristics	16
Inter-rater reliability	16
Classification agreement.....	16
Signal quality	17
3.2 <i>Electrode reduction for visual assessment</i>	17
Patient characteristics	17
Inter-rater reliability	18
Classification agreement.....	19
Outcome prediction	19
3.3 <i>Electrode reduction for qEEG analysis</i>	20
Patient characteristics	20
Predetermined reduced sets	20
L1 regularization.....	22
4 Discussion	23
4.1 <i>BrainStatus</i>	23
4.2 <i>Electrode reduction for visual assessment</i>	24
4.3 <i>Electrode reduction for qEEG analysis</i>	25
5 Recommendations	27
6 Conclusion	27
References	28
Appendices	31
Appendix A.1 <i>BrainStatus</i>	31
Appendix A.2. <i>User interface scoring application</i>	32
Appendix A.3. <i>Additional results BrainStatus</i>	33
Appendix A.4. <i>Artefact scores BrainStatus</i>	36
Appendix A.5. <i>Additional results electrode reduction for visual assessment</i>	37
Appendix A.6. <i>ROC curves reduced montages</i>	38
Appendix A.7. <i>Performance LR model reduced montages</i>	39
Appendix A.8. <i>Boxplots feature values</i>	41
Appendix A.9. <i>Performance and trained weights L1 regularization</i>	45
Appendix A.10. <i>LR performance compared with literature</i>	46

List of figures

Figure 1: Schematic overview of electrode positions and used EEG montages	11
Figure 2: Confusion matrix EEG background patterns: BrainStatus vs bipolar montage	17
Figure 3: Confusion matrix EEG background patterns: frontotemporal vs bipolar montage	18
Figure 4: ROC curve all channels bipolar montage	21
Figure 5: ROC curve all channels refCz montage	21
Figure 6: ROC curves frontotemporal vs bipolar montage	22
Figure A.1: Overview of the BrainStatus	31
Figure A.2: User interface of visual scoring web application	32
Figure A.3: Confusion matrix EEG background patterns: BrainStatus vs bipolar montage	33
Figure A.4: Confusion matrix EEG background patterns: frontotemporal vs BrainStatus montage	33
Figure A.5: Confusion matrix EEG background patterns: frontotemporal vs bipolar montage	34
Figure A.6: Confusion matrix EEG rhythmic patterns: BrainStatus vs bipolar montage	34
Figure A.7: Confusion matrix EEG rhythmic patterns: frontotemporal vs BrainStatus montage	34
Figure A.8: Confusion matrix EEG rhythmic patterns: frontotemporal vs bipolar montage	34
Figure A.9: Confusion matrix assigned outcome: BrainStatus vs bipolar montage	35
Figure A.10: Confusion matrix assigned outcome: frontotemporal vs BrainStatus montage	35
Figure A.11: Confusion matrix assigned outcome: frontotemporal vs bipolar montage	35
Figure A.12: Plotted artefact scores over time for all patients	36
Figure A.13: Confusion matrix EEG rhythmic patterns: frontotemporal vs bipolar montage	37
Figure A.14: Confusion matrix assigned outcome: frontotemporal vs bipolar montage	37
Figure A.15: ROC curves for all reduced montages compared to the full montages	38
Figure A.16: Boxplots of the feature values at 12 hours after CA for all patients	42
Figure A.17: Boxplots of the feature values at 24 hours after CA for all patients	44

List of tables

Table 1: Extracted qEEG features and descriptions	15
Table 2: Patient and cEEG characteristics BrainStatus	16
Table 3: Patient characteristics FT montage	18
Table 4: 2x2 table prediction of poor outcome: bipolar vs frontotemporal montage	19
Table 5: 2x2 table prediction of good outcome: bipolar vs frontotemporal montage	19
Table 6: Sensitivity and specificity for outcome prediction: bipolar vs frontotemporal montage	19
Table 7: Patient characteristics qEEG analysis	20
Table A.1: ICC values BrainStatus, frontotemporal and bipolar montage	33
Table A.2: Classification agreement between BrainStatus, frontotemporal and bipolar montage	35
Table A.3: ICC values frontotemporal and bipolar montage	37
Table A.4: Classification agreement between frontotemporal and bipolar montage	37
Table A.5: Performance LR model at 12h after CA: reduced montages vs full bipolar montage	39
Table A.6: Performance LR model at 24h after CA: reduced montages vs full bipolar montage	39
Table A.7: Performance LR model at 12h after CA: reduced montages vs full refCz montage	39
Table A.8: Performance LR model at 24h after CA: reduced montages vs full refCz montage	39
Table A.9: Performance LR model at 12h after CA: frontotemporal vs bipolar montage	40
Table A.10: Performance LR model at 24h after CA: frontotemporal vs bipolar montage	40
Table A.11: Trained weight for all channels with L1 regularization at 12h after CA	45
Table A.12: Corresponding model performance using L1 regularization at 12h after CA	45
Table A.13: Trained weight for all channels with L1 regularization at 24h after CA	45
Table A.14: Corresponding model performance using L1 regularization at 24h after CA	45
Table A.15: Performance of earlier developed outcome prediction models at 12h after CA	46
Table A.16: Performance of earlier developed outcome prediction models at 24h after CA	46

1 Introduction

Out-of-hospital cardiac arrest (OHCA) is a serious health issue and a major cause of unexpected death in developed countries, with survival rates ranging from less than 5% to 35%.¹ Patients who achieve return of spontaneous circulation (ROSC) after OHCA are usually admitted to the Intensive Care Unit (ICU) and treated with targeted temperature management (TTM), including standardized sedation to prevent shivering that might lead to warming.¹⁻³ Due to cessation of blood supply to the brain during CA, these patients often suffer from hypoxic-ischemic brain injury and postanoxic coma. Approximately 176,000 patients with postanoxic coma after CA are admitted to the ICU yearly in Europe.⁴ Of these patients, 40-66% never regain consciousness as a result of severe postanoxic encephalopathy.⁴ In case a patient does not wake up after TTM is terminated and sedation is discontinued, the decision has to be made between continuation or withdrawal of life-sustaining treatment. Early prediction of recovery perspectives may guide these decisions. Identification of patients with poor neurological outcome in an early stage can prevent continuation of futile medical treatment, decrease ICU stay and medical costs, and shorten the time of uncertainty for the patient's family.^{5,6}

Continuous electroencephalography (cEEG) is increasingly used for prognostication after cardiac arrest, and several studies report that the first 24 hours of EEG after resuscitation have the highest prognostic value for neurological outcome.⁶⁻⁹ Hofmeijer et al.¹⁰ showed that rapid recovery towards continuous EEG patterns within 12 hours was almost invariably associated with a good neurological outcome. These results emphasize the importance of initiation of EEG monitoring as soon as possible after ROSC. However, nowadays in clinical practice, the ICU staff is not trained in recording and reviewing EEG.^{11,12} The role of technicians and clinical neurophysiologists is essential for successful EEG monitoring, but they are usually not based in the ICU, resulting in shortage of 24/7 availability of specialized staff.^{13,14} Furthermore, the electrode locations have to be determined very accurately, which is time-consuming. These factors can cause delayed start of monitoring, resulting in missing useful information from registration shortly after ROSC.

In addition to the issues regarding the application of the electrodes, the EEG signals obtained from continuous monitoring are currently visually analyzed to interpret the clinical value. This visual assessment has to be performed by trained and experienced specialists and can be very time-consuming and partially subjective.¹⁵⁻¹⁸ Quantitative EEG measures can be a helpful solution for these issues.^{15,19} Tjepkema-Cloostermans et al. composed a Cerebral Recovery Index (CRI) based on extraction of specific features from the EEG signal, and several revisions of the algorithm have followed.^{8,9,20} These studies show promising results with high prognostic value for both poor and good neurological outcome, with highest performance at 9h after CA with a sensitivity of 78% at 100% specificity and 63% sensitivity at 95% specificity, respectively.²⁰ Also, logistic regression models are being developed, and interest in using neural networks for EEG analysis is growing and showing potentially useful predictions.²¹⁻²⁴ However, these studies all used data obtained from 9 or even 21 EEG electrodes. Interest in using simplified, reduced montages for qEEG analysis is growing to improve the ease and speed of lead application in clinical practice.²⁵

The extensive procedure of electrode application introduces the need for EEG solutions that are fast and easy to set up without extensive training. One of the clinically available, easy-to-use options is the BrainStatus, developed by Lepola et al.^{26,27} The BrainStatus consists of 16 electrodes, all embedded into a flexible polyester film and coated with an adhesive hydrogel membrane. An overview of the BrainStatus is given in Appendix A.1. Lepola et al. already showed promising results by testing the BrainStatus for ruling out status epilepticus in a low number of patients.^{28,29} To explore if the BrainStatus would also be suitable for cEEG monitoring, several explorative studies have been performed in the Amsterdam UMC.^{30–32} Unfortunately, the first results showed a substantial presence of artefacts in the EEG signals measured with the BrainStatus, presumed to be caused by increasing impedance due to loosening of the electrodes³², impeding reliable analysis of the EEG pattern. A possible solution might be to perform intermittent check-ups and thereby decrease the number of artefacts and improve the signal quality.

Furthermore, an explorative study is currently running at the AMC for testing the ability of alternative, adhesive electrodes from other domains (ECG, EMG) to register EEG signals with sufficient signal quality. Preliminary results show acceptable impedances, for example for the Ambu® Neuroline 720, resulting in signal quality comparable to the conventional cup electrodes, raising the question if these electrodes could also have potential to be used for cEEG monitoring after CA at the ICU.³³

Alternative monitoring methods with adhesive electrodes are inseparably linked to limited possible locations for electrode placement, resulting in reduced electrode montages. Several studies have been conducted to find the efficacy of reduced electrode sets for prediction of neurological outcome using qEEG analysis. The results offer potential in developing a clinically practical method for cEEG monitoring.^{25,34–37} An earlier study in the AMC compared the performance of a CRI algorithm on a bipolar montage consisting of 9 electrodes and a central montage based on 3 electrodes and reported better performance for the central montage.³⁸ Although the first results are promising, validation of the predictive value of reduced electrode sets is required for both visual assessment and qEEG analysis.

During this study, we aimed to optimize both the monitoring and analysis methods for using cEEG in the ICU as a predictive tool for neurological outcome after CA. The main objective was to explore possibilities for simplifying the application procedure while preserving the predictive value of the EEG signal for both visual assessment and qEEG analysis. Three components of the study can be distinguished. First, the BrainStatus electrode was used to investigate if its registration quality improves with intermittent check-ups. Second, the classification agreement and predictive value of a 4-channel frontotemporal montage compared to the conventional 9-channel bipolar montage were examined to validate the use of a reduced sub-hairline montage for visual assessment. Lastly, we aimed to answer the question whether reduced montages can be used for reliable prognostication after CA using qEEG analysis, with minimal decrease of performance compared to full montages.

2 Materials and methods

Study population

Continuous EEG was obtained from postanoxic comatose patients after CA admitted to the ICU of Amsterdam UMC – location AMC. Patients were eligible for inclusion if cEEG monitoring was started within 24 hours after CA. Exclusion criteria included traumatic brain injury, acute stroke, progressive neurodegenerative disease, prearrest modified Rankin scale ≥ 4 or prearrest life expectancy ≤ 6 months based on comorbidity. The exact study population was different for the three subparts of the study, described in corresponding subsections.

Outcome assessment

Neurological outcome was used as the primary outcome measure for all parts of the study, defined as a score on the Cerebral Performance Category (CPC) scale. Outcome was dichotomized into good outcome for CPC 1-2 (indicating no or mild neurological impairment) and poor outcome for CPC 3-5 (indicating severe neurological impairment vegetative state or death).^{39,40} Also, several EEG patterns at specific timepoints are associated with either poor or good neurological outcome, according to the Dutch guidelines for prognosis of postanoxic coma.⁴¹ Patterns known as indicative for a poor neurological outcome are a suppressed background pattern, burst suppression with identical bursts or a continuous background pattern with low voltage at 24 hours after CA. Continuous background patterns with normal voltage at 12 hours after CA are categorized as indicative for good outcome.^{6,8,39,41}

Standard of care and monitoring

Conventional monitoring of cEEG on the ICU in the AMC is done with eleven Ag/AgCl cup electrodes and was registered for all patients in all parts of the study. Electrodes are being placed according to the standardized 10-20 system at positions Fp1, Fp2, T3, T4, C3, C4, Cz, O1 and O2, and the ground and reference electrode are placed in mid-line. Recording is started as soon as possible after the start of TTM and is continued up to three days unless the patient regains consciousness or dies. As soon as the patient is admitted to the ICU, TTM is started at 36 °C and maintained for 24 hours. During TTM, propofol is administered with a maximum of 5 mg/kg/hour to keep the patient sedated to a Richmond Agitation-Sedation Scale of -4.⁴²

Epoch selection and EEG montages

Epoch selection and preprocessing were done using Matlab R2020b. We selected 5-minute epochs from the cEEG data from a 20-minute window around specified timepoints. Selection was based on the least number of artefacts, using an earlier developed algorithm.⁹ The number of artefacts is based on the number of high voltage peaks, the power frequency ratio between the EEG range and higher frequencies, and the number of channels containing zeros. Epochs were selected using this algorithm around 12 and 24 hours after CA if available for all patients in all parts of the study. For patients monitored with the BrainStatus, epochs were selected around every whole hour available after CA.

All EEG signals were re-referenced into a specific montage before analysis. Three different EEG montages were used for the different parts of the study: a longitudinal bipolar montage, a frontotemporal montage, and a montage where all electrodes are referenced to Cz. An overview of the electrode positions of all used montages is shown in Figure 1. We used the Cz montage to create a single channel for each electrode position while keeping comparable distances between the electrode and Cz used as reference for all channels. The montage for EEG signals registered with the BrainStatus is based on the electrode positions and channels from the frontotemporal montage.

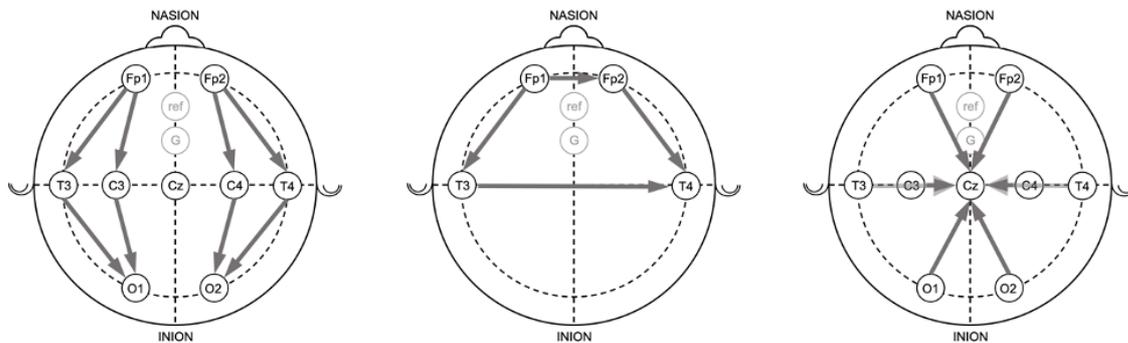


Figure 1: Schematic overview of electrode positions and used EEG montages: longitudinal bipolar montage (left), frontotemporal montage (middle) and all electrodes referenced to Cz (right). Electrode locations used from the BrainStatus are based on the locations of the frontotemporal montage.

Visual scoring application

The EEG epochs to be visually scored were presented to three experienced EEG readers (AFvR, JH, MMA) individually using a custom-built Matlab application, built using Matlab R2020b. Within the app, the signals were preprocessed by applying an adjustable third-order Butterworth band-pass filter (high-pass cut-off 0.300, 0.530, 1, 1.6, 5.3 or 10 Hz, low-pass 2, 10, 15, 20, 30, 50, 60, 70 or 120 Hz) and a switchable notch filter (band-stop for 49-51 Hz). Additionally, all EEG samples were resampled to 256 Hz and re-referenced into a longitudinal bipolar montage, a frontotemporal montage and if applicable into a BrainStatus montage. The observers were blinded for patients' condition, outcome, electrode sets and the time after cardiac arrest. All EEG samples were scored on background pattern and rhythmic pattern individually by all observers with the help of systematic drop-down menus. Scoring criteria were based on Hirsch' standardized terminology.⁴³ Background pattern could be scored as suppressed, burst-suppression with or without identical bursts, discontinuous or continuous. Independently, rhythmic pattern was scored as spike-wave pattern, periodic discharges, rhythmic delta, or none. For the user interface and additional information on the drop-down menus and scoring criteria, see Appendix A.2. Majority vote was used to determine the final classification of the epoch for all characteristics. When there was no majority vote, a consensus meeting was planned to reach agreement between the observers. This resulted in one classification for each EEG epoch. The performance was evaluated by calculating the inter-rater reliability and the classification agreement between either the BrainStatus montage or the frontotemporal montage and the 9-channel bipolar montage as the golden standard. As prognostication is mainly based on the background pattern, this was the main characteristic to evaluate.

Inter-rater reliability

We calculated the inter-rater variability using the intra-class correlation coefficient (ICC) and its 95% confidence interval, based on an average of independent measurements, absolute-agreement, two-way random-effects model. According to the value of the ICC, a distinction was made between poor (< 0.50), moderate (0.50 – 0.75), good (0.75 – 0.90) or excellent (> 0.90) inter-rater reliability.⁴⁴

Classification agreement

The classifications of all EEG epochs were plotted in confusion matrices and Cohen's kappa was calculated to evaluate the agreement in classification between the different montages. Labels were assigned to certain ranges of the value of kappa to indicate the strength of agreement, according to the benchmarks of Landis and Koch.⁴⁵ We distinguished between poor (< 0), slight (0 – 0.20), fair (0.21 – 0.40), moderate (0.41-0.60), substantial (0.61 – 0.80) and (almost) perfect (0.81 – 1.00) agreement. The agreement between two montages was defined as reliable if the strength of agreement was at least substantial, so a value for kappa of at least 0.61.

2.1 BrainStatus

Study population and outcome assessment

Data for the BrainStatus analysis was included prospectively on the ICU in the AMC, with additional exclusion criteria being proved or suspected COVID-19 infection, treatment with an extracorporeal membrane oxygenation (ECMO) machine or planned transfer to another hospital within one day. Due to the prospective inclusion, neurological outcome at 6 months after CA was not yet available, so we only assigned an outcome prediction to all EEG samples based on the visual scoring.

Monitoring and check-ups

Monitoring with the BrainStatus was always done simultaneously with the conventional EEG monitoring. Overlapping electrodes (Fp1 and Fp2) were placed as close as possible to their intended location. To investigate if the artefacts found in the EEG signals registered with the BrainStatus result from increasing impedance and loosening of electrodes, intermittent check-ups were performed. A check-up was done every 2,5 to 3 hours from the start of registration with the BrainStatus, during office hours. During the check-ups, we pressed the BrainStatus firmly upon the skin and checked and saved the impedance just before and directly after doing this. The signals registered with the BrainStatus electrodes are evaluated based on the agreement in visual scoring and the signal quality over time.

Visual scoring

The EEG signal of all patients monitored with the BrainStatus was visually scored in three different montages: the longitudinal bipolar montage, the frontotemporal montage and the BrainStatus montage. The performance was evaluated by calculation of the inter-rater variability and classification agreement between the different montages.

Evaluation of signal quality

To evaluate the signal quality over time, we extracted 5-minute epochs from the EEG signal around every whole hour after the time of cardiac arrest. Signals from all single channels were extracted from the file and were resampled to 256 Hz. We calculated an artefact score for all epochs for all channels of the conventional cup electrodes and all channels of the BrainStatus, based on the number of high amplitude peaks and the power frequency ratio. All epochs are split in fragments of 30 seconds for calculation. Before calculating the number of high amplitude peaks, we applied a third-order Butterworth band-pass filter between 0.5 and 30 Hz. Peaks are assigned as high amplitude if the amplitude is above the mean plus two times the standard deviation of that fragment. The final amplitude score is the number of times the amplitude exceeds the threshold divided by the total length of the signal. Because higher frequencies are essential for the calculation of the power frequency ratio, a third-order Butterworth high-pass filter of 0.5 Hz was applied instead of a band-pass filter. We calculated the power frequency ratio between the EEG frequency band (0.5-25 Hz) and the EMG frequency band (25-40 Hz). The final frequency artefact score was determined as EMG power divided by EEG power. Both scores resulted in values between 0 and 1. The scores were averaged over the time fragments and subsequently summed to obtain one artefact score per channel and per epoch.

2.2 Electrode reduction for visual assessment

Study population and outcome assessment

Data used for visual assessment of the frontotemporal montage was based on the dataset from an earlier outcome prediction study from Admiraal et al.⁴⁶ Data was included between April 2015 and February 2018. The assigned outcome prediction based on visual scoring and the neurological outcome at 6 months after CA were both used for this analysis.

Visual scoring

To validate the predictive value of the frontotemporal montage, we reassessed 221 EEG epochs that were visually scored in a bipolar montage in Admiraal's reactivity study.⁴⁶ The same EEG epochs were assessed in the frontotemporal montage by the same three observers (AFvR, JH, MMA), using the visual scoring application.

Outcome prediction

In addition to the calculation of the inter-rater reliability and the classification agreement between the two montages, we compared the sensitivity and specificity for prediction of neurological outcome between the bipolar montage and the frontotemporal montage. The predictive value of patterns indicative for both poor and good neurological outcome was evaluated in both montages and compared using a McNemar test.

2.3 Electrode reduction for qEEG analysis

Study population and outcome assessment

The dataset for qEEG analysis was based on the dataset from another earlier outcome prediction study in the AMC²⁴ and was included between May 2014 and January 2020. Neurological outcome at 6 months after CA was assessed for all patients.

Preprocessing and model input

Before feature extraction, we filtered the EEG signal with a third-order Butterworth band-pass filter between 0.5 and 30 Hz and resampled the signal to 128 Hz. Subsequently, we re-referenced the EEG signal into the longitudinal bipolar montage and the Cz montage, and both montages were separately used for feature extraction. Testing the possibilities of reduced electrode sets for prediction of neurological outcome was based on an earlier developed algorithm using logistic regression (LR).²⁴ We extracted features from the EEG signal to use as the input for the model. Neurological outcome at 6 months after CA was used as output parameter, where a value of 1 represents poor outcome and a value equal to 0 represents good neurological outcome.

Feature extraction

The preprocessed 5-minute EEG epochs were segmented into 30 non-overlapping 10-second time fragments. A total of 12 features was extracted from the EEG signal, calculated for all channels and time fragments separately. After calculation, all features were averaged over all used channels and time fragments. All features were normalized and scaled between 0 and 1 with respect to the same feature of all patients. Extracted features were in line with previous studies in the AMC and based on extensive literature study and multicollinearity analysis.^{24,47} Table 1 summarizes the extracted features, categorized into three domains.

Logistic regression model

The LR model was built using Pytorch 1.6.0 in Python 3.7.10. Within the model, we used a sigmoid activation function to obtain the probability of the outcome being equal to 1. The dataset was randomly divided into 10-folds for cross-validation. Another 10% of the data was used for validation during training, resulting in 80% actual training set, 10% validation set and 10% test set. The validation set was used to implement learning rate decay with early stopping for optimal performance. The initial learning rate was set at α equal to 0.01 and after decay the learning rate α was 0.0001. To obtain a reliable and robust estimation of the model performance, we repeated the 10-fold cross-validation 50 times.

Performance metrics

In line with earlier reported outcome prediction studies, the model's performance was evaluated using three metrics retrieved from the receiver operating characteristic (ROC) curve: area under the curve (AUC) and sensitivities for predefined specificity thresholds for good and poor outcome. To avoid withdrawal of life-sustaining treatment in patients with a viable outcome, it is imperative to predict poor outcome at a zero false positive rate. Therefore, we used sensitivity at 100% specificity (Se100) for prediction of poor outcome. For good outcome, we retrieved the sensitivity at 95% specificity (Se95).

Table 1: Extracted qEEG features and descriptions. Table adapted from Van Poppe²⁴.

Domain		Feature	Description
Complexity	1	Tsallis entropy ^{48,49}	A measure to quantify the uncertainty of a stochastic signal
	2	False nearest neighbour (FNN) ⁵⁰	A measure to quantify the degree of stochasticity in a signal by estimating the embedding dimension, indicating its constancy and smoothness
	3	Autoregressive (AR) coefficient $2^{51,52}$	Estimated nonseasonal autoregressive term coefficient at t-2 of a second-order autoregressive model given the EEG signal
Category	4	Normalized theta power ⁵³⁻⁵⁵	Theta power (4-7 Hz) divided by the total power (0.5-30 Hz)
	5	Normalized alpha power ⁵³⁻⁵⁵	Alpha power (8-13 Hz) divided by the total power (0.5-30 Hz)
	6	Normalized beta power ⁵³⁻⁵⁵	Beta power (14-30 Hz) divided by the total power (0.5-30 Hz)
	7	Signal power ⁵³⁻⁵⁵	The total power in the frequency range of interest (0.5-30 Hz)
	8	Regularity ⁹	A measure to quantify regularity in amplitude of the signal
	9	Number of epileptic spikes ⁵⁶	The number of epileptiform spikes in the EEG
	10	Burst suppression ratio (BSR) ²⁰	The ratio of the duration of the EEG signal with an amplitude equal or lower than 5 μ V to the duration of the entire signal
Connectivity	11	Delta coherence ⁹	A measure to quantify the degree of similarity in the delta band
	12	Phase lag index (PLI) ⁵⁷	A measure to quantify phase synchronization, indicating the level of asymmetry between two signals

Predetermined reduced sets

To test the performance of the LR model with input features based on a reduced number of electrodes, we predefined reduced electrode sets. We compared these model performances with the performance on the full montage. We created montages consisting of the two adjacent channels of one electrode from the bipolar montage (e.g. Fp1-T3 and Fp1-C3 for Fp1), and single channels were used from the Cz reference montage (e.g. Fp1-Cz). The two connectivity features (delta coherence and phase lag index) are calculated for all combinations of channels within a montage, so were assumed unreliable in these reduced montages and left out as input features. Additionally, we created a reduced montage according to the electrode locations from the frontotemporal montage (Fp1-Fp2, Fp1-T3, Fp2-T4). The differences between the model's performance metrics were compared using an unpaired two-sample students t-test. The difference was considered statistically significant if $p < 0.05$.

L1 regularization

To see if the LR model was able to identify which electrode channels contributed most to the outcome prediction within the model, we applied L1 regularization, which pushes the weights of less essential features towards zero. To distinguish between electrode channels, isolation of the channels is required to obtain a single feature value per channel. First, a change in feature value must resemble a similar outcome change for all features. When necessary, normalized features were inverted, resulting in a higher feature value indicating a higher probability for poor outcome. Which features to invert was based on the median value of the feature for all subjects. Subsequently, we summed all features per channel, resulting in one feature value per channel. We determined the average of the absolute weights over all repetitions of the LR model for each channel, alongside performance tracking of the model, while applying different values of the regularization factor λ .

3 Results

The results of the study are separated into three subparts, including the BrainStatus study, the electrode reduction for visual assessment and electrode reduction for qEEG analysis. All parts of the study have separate study populations and patient characteristics.

3.1 BrainStatus

Patient characteristics

Table 2 shows the characteristics of included patients in the BrainStatus study with additional check-ups. A total of six patients was included from October 2020 till April 2021. EEG was monitored at 12 hours after CA for three patients and the timepoint 24 hours after CA was available for four patients. EEG patterns were extracted from the patient file.

Table 2: Patient and cEEG characteristics BrainStatus

	Monitored with Brainstatus (n = 6)
Patient characteristics	
Age	75 [60-79]
Male	3/6 (50%)
OHCA	4/6 (67%)
Witnessed arrest	6/6 (100%)
Time to ROSC (min)	21 [12-38]
Initial rhythm shockable	4/6 (67%)
Cardiac etiology	5/5 (100%)
Died at ICU	2/6 (33%)
cEEG characteristics	
EEG pattern at 12h after CA (n = 3)	
BS without identical bursts	2
Discontinuous	1
EEG pattern at 24h after CA (n = 4)	
BS without identical bursts	1
Discontinuous	2
Continuous	1

Data presented as median [interquartile range] or n/total (%). OHCA: out of hospital cardiac arrest; ROSC: return of spontaneous circulation; EEG: electroencephalography; CA: cardiac arrest; BS: burst-suppression

Inter-rater reliability

The inter-rater variability for scoring background pattern was poor for the BrainStatus montage (ICC = 0.07, 95% CI: -3.24-0.84) and good in the bipolar montage (ICC = 0.82, 95% CI: 0.37-0.97). This indicates a lower level of agreement between the observers when scoring the epochs in the BrainStatus montage compared to the bipolar montage. Values for the ICC for scoring rhythmic pattern and the assigned outcome prediction can be found in Appendix A.3 for all montages.

Classification agreement

The confusion matrix for the scored background pattern in the BrainStatus montage compared to the bipolar montage is shown in Figure 2. The classification within the BrainStatus montage disagreed with the bipolar montage on two EEG epochs. The corresponding value for Cohen's kappa equals 0.48 (95%

CI: -0.13-1.09), indicating a moderate agreement. Confusion matrices and kappa values for other montage combinations and characteristics can be found in Appendix A.3. It should be noted that the patterns scored using the visual scoring app do not entirely agree with the assessed EEG patterns extracted from the patient files as shown in Table 2.

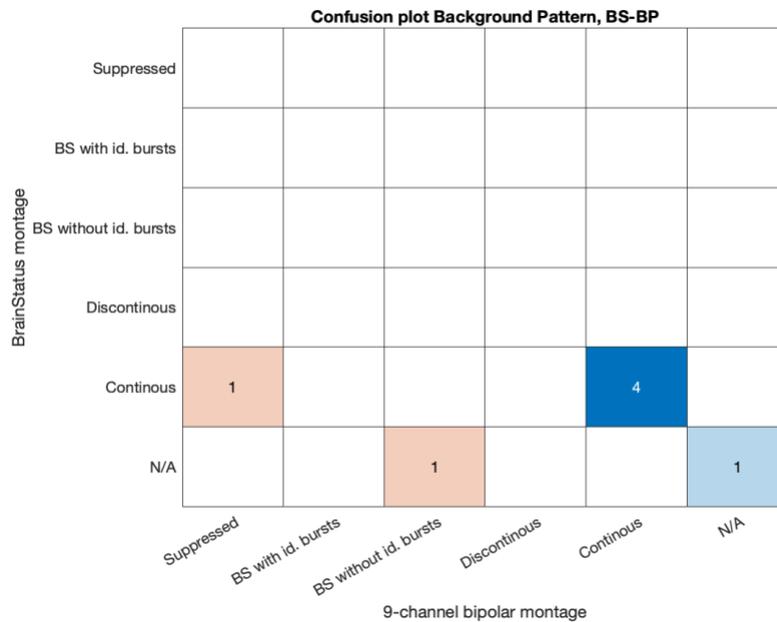


Figure 2: Confusion matrix for scored EEG background patterns comparing the BrainStatus montage with the bipolar montage. N/A is not assessable, blue colour indicates conformity and red represents disagreement. Colour saturation corresponds with the relative number of observations. The classification in the BrainStatus montage disagrees with the classification in the bipolar montage for two EEG epochs.

Signal quality

The artefact score was determined for all available timepoints for all included patients. If the EEG recording was paused or too many artefacts were present in the signal, no epoch could be extracted for that timepoint. For two patients, the artefact score for the Brainstatus increases clearly over time and increases more rapidly than the conventional cup electrodes. For two other patients, the mean artefact score is slightly higher for the cup electrodes than for the BrainStatus electrodes. And for the last two patients, the mean artefact scores for the cup electrodes and the BrainStatus are nearly equal. No evident decreases or other changes are seen directly after the check-ups. The artefact scores over time for all patients separately can be found in Appendix A.4

3.2 Electrode reduction for visual assessment

Patient characteristics

An existing dataset of 154 patients was used, with characteristics as described in Table 3. For 69 of these patients, EEG at 12 hours after CA was available and EEG at 24 hours after CA was available for almost all (152) patients. A distinction is made between patients with good neurological outcome and patients with poor outcome.

Table 3: Patient characteristics FT montage

	Good outcome (n = 78)	Poor outcome (n = 76)
Patient characteristics		
Age	62 [51-70]	65 [52-73]
Male	65/78 (83%)	53/76 (70%)
OHCA	72/78 (92%)	65/76 (86%)
Witnessed arrest	62/76 (82%)	50/72 (69%)
Time to ROSC (min)	13 [10-18] ^a	22 [15-32] ^b
Initial rhythm shockable	69/74 (93%)	37/73 (51%)
Cardiac etiology	63/71 (89%)	35/66 (53%)
cEEG characteristics		
EEG pattern at 12h after CA (n = 69)		
Suppressed	0/39 (0%)	11/30 (37%)
BS with identical bursts	0/39 (0%)	2/30 (7%)
BS without identical bursts	0/39 (0%)	7/30 (23%)
Discontinuous	13/39 (33%)	5/30 (17%)
Continuous	25/39 (64%)	4/30 (13%)
Unknown	1/39 (3%)	1/30 (3%)
EEG pattern at 24h after CA (n = 152)		
Suppressed	0/78 (0%)	20/74 (27%)
BS with identical bursts	0/78 (0%)	5/74 (7%)
BS without identical bursts	2/78 (3%)	10/74 (14%)
Discontinuous	9/78 (12%)	10/74 (14%)
Continuous	66/78 (85%)	26/74 (35%)
Unknown	1/78 (1%)	3/74 (4%)

Data presented as median [interquartile range] or n/total (%). OHCA: out of hospital cardiac arrest; ROSC: return of spontaneous circulation; EEG: electroencephalography; CA: cardiac arrest; BS: burst-suppression. ^an = 75, ^bn = 69.

Inter-rater reliability

The calculated ICC for scoring background pattern equals 0.94 for both montages with a 95% confidence interval of 0.92-0.95 in the frontotemporal montage and 0.93-0.95 in the bipolar montage, indicating excellent agreement between the three observers. Values for the ICC for scoring rhythmic pattern and the assigned outcome prediction can be found in Appendix A.5 for both montages.

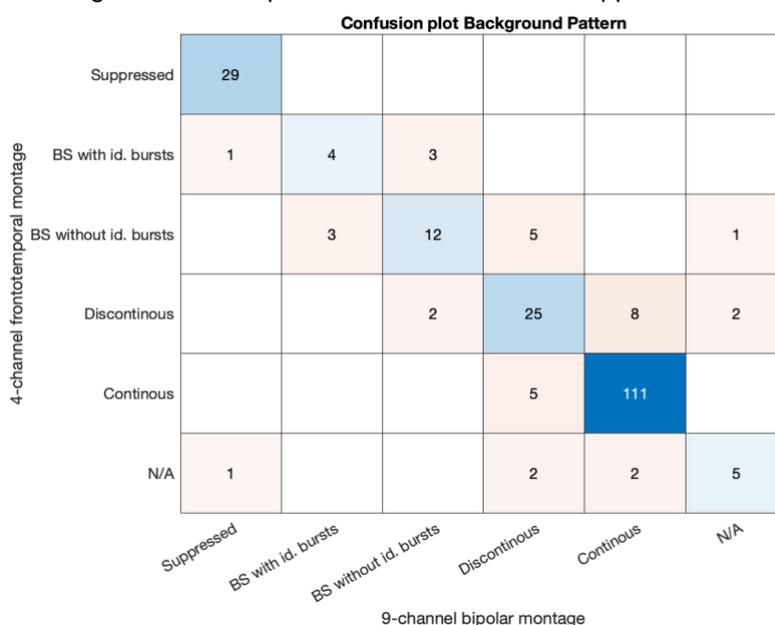


Figure 3: Confusion matrix for scored EEG background patterns comparing the 4-channel frontotemporal montage with the 9-channel bipolar montage. N/A is not assessable, blue colour indicates conformity and red represents disagreement. Colour saturation corresponds with the relative number of observations. For 186 of 221 EEG epochs the scored background pattern agreed between the two montages.

Classification agreement

The classification agreement between the two montages for scoring background pattern is shown in the confusion matrix in Figure 3. The corresponding value for kappa equals 0.76 (95% CI: 0.69-0.83), indicating substantial agreement between the two montages. The confusion matrices and kappa values for rhythmic pattern and assigned outcome prediction can be found in Appendix A.5.

Outcome prediction

The 2x2 tables for prediction of poor outcome at 24 hours after CA and prediction of good outcome at 12 hours after CA can be found in Table 4 and Table 5. Sensitivities and specificities for both good and poor outcome are shown in Table 6. Poor outcome was predicted with 100% specificity in both montages and 34% and 31% sensitivity in respectively the bipolar montage and frontotemporal montage. Prediction of good outcome based on continuous EEG pattern at 12 hours after CA could be done with a specificity of 87% for both montages and sensitivities of 64% and 59% for the bipolar montage and the frontotemporal montage, respectively. The McNemar test showed no statistical difference between the sensitivities and specificities of the bipolar and frontotemporal montage for both good and poor neurological outcome.

Table 4: 2x2 table for the prediction of poor outcome at 24h after CA based on 9-channel bipolar montage and 4-channel frontotemporal montage.

	Poor outcome	Good Outcome
Background pattern scored		
Bipolar montage		
Suppressed, BS with identical bursts or low voltage	25	0
Other	49	78
Frontotemporal montage		
Suppressed, BS with identical bursts or low voltage	23	0
Other	51	78

BS: burst-suppression

Table 5: 2x2 table for the prediction of good outcome at 12h after CA based on 9-channel bipolar montage and 4-channel frontotemporal montage.

	Good outcome	Poor outcome
Background pattern scored		
Bipolar montage		
Continuous (normal voltage)	25	4
Other	14	26
Frontotemporal montage		
Continuous (normal voltage)	23	4
Other	16	26

Table 6: Sensitivity and specificity for prediction of neurological outcome based on EEG background pattern using a 9-channel bipolar montage and a 4-channel frontotemporal montage.

EEG pattern	Time after CA	Predicting	Montage	Sens (95% CI)	Spec (95% CI)
Continuous (normal voltage)	12 h	Good outcome	BP	64 (47-79)	87 (69-96)
			FT	59 (42-74)	87 (69-96)
Suppressed, BS with identical bursts or low voltage	24 h	Poor outcome	BP	34 (23-46)	100 (95-100)
			FT	31 (21-43)	100 (95-100)

EEG: electroencephalography; BS: burst-suppression; CA: cardiac arrest; Sens: sensitivity; Spec: specificity; CI: confidence interval

3.3 Electrode reduction for qEEG analysis

Patient characteristics

An existing dataset of 186 patients monitored between May 2014 and January 2020 was used, with characteristics described in Table 7. For 78 of these patients, EEG at 12 hours after CA was available and EEG at 24 hours after CA was available for 181 patients. A distinction is made between patients with good neurological outcome and patients with poor outcome.

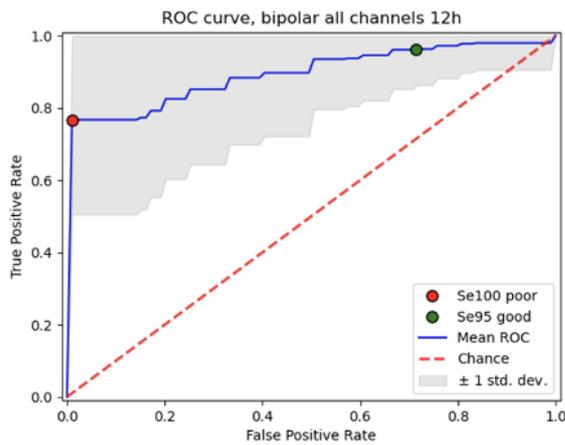
Table 7: Patient characteristics qEEG analysis

	Good outcome (n = 76)	Poor outcome (n = 110)
Patient characteristics		
Age	61 [51-71]	64 [53-72]
Male	59/76 (78%)	80/110 (73%)
OHCA	69/76 (91%)	92/110 (84%)
Witnessed arrest	58/67 (87%)	62/91 (68%)
Time to ROSC (min)	13 [9-18.5] ^a	23 [15-37] ^b
Initial rhythm shockable	67/75 (89%)	53/106 (50%)
Cardiac etiology	59/64 (92%)	53/93 (57%)
cEEG characteristics		
EEG available at 12h after CA	41/76 (54%)	37/110 (34%)
EEG available at 24h after CA	74/76 (97%)	107/110 (97%)

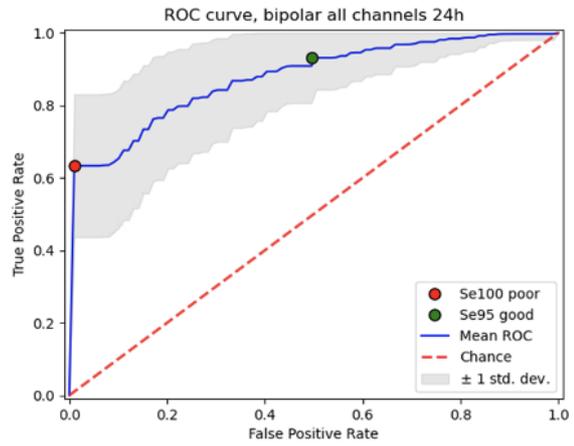
Data presented as median [interquartile range] or n/total (%). OHCA: out of hospital cardiac arrest; ROSC: return of spontaneous circulation; EEG: electroencephalography; CA: cardiac arrest.
^an = 73, ^bn = 91.

Predetermined reduced sets

The input features averaged over all channels is used as the baseline input to compare with. All results are presented as mean (95% CI). The full bipolar montage shows an AUC of 0.897 (0.885-0.908) with a sensitivity of 77% (74-79) at 100% specificity for predicting poor outcome and 71% (68-74) sensitivity at 95% specificity for prediction of good outcome at 12 hours after CA. Its corresponding ROC curve with its standard deviation can be found in Figure 4a. For 24 hours after CA, the AUC equals 0.879 (0.873-0.886), with a sensitivity of 63% (62-65) at 100% specificity for predicting poor outcome and 50% (47-52) sensitivity at 95% specificity for prediction of good outcome. The corresponding ROC curve is shown in Figure 4b. ROC curves for the reduced montages can be found in Appendix A.6. Eliminating the connectivity features from the input doesn't affect the model's performance at 12 hours after CA, but the Se₉₅ at 24 hours after CA is significantly decreasing. Furthermore, the model's performance is not negatively affected by reducing the input channels at 12 hours after CA. The channels adjacent to T4 even show significantly higher values for the AUC and Se₉₅. At 24 hours after CA, the performance increases for the channels related to T3 and C3, and some metrics decrease for the channels related to Fp2 and O2. Performance metrics with 95% confidence intervals and their corresponding p-values for all reduced montages compared to the full bipolar montage at 12 hours and 24 hours after CA can be found in Appendix A.7.



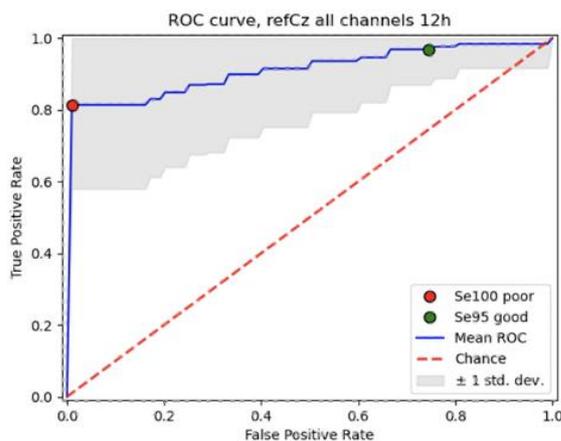
a) ROC curve for averaged input over all channels in the bipolar montage at 12h after CA



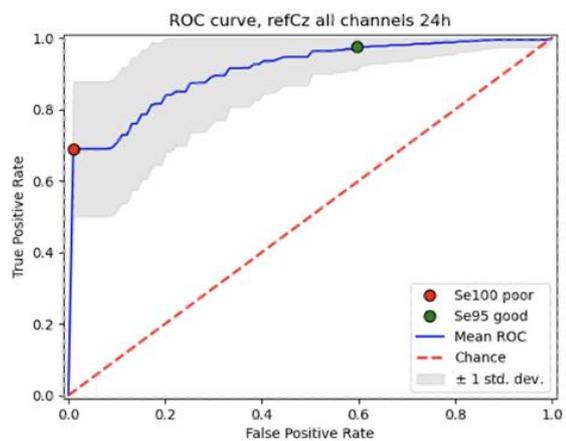
b) ROC curve for averaged input over all channels in the bipolar montage at 24h after CA

Figure 4: Mean ROC curve and its standard deviation over 50 iterations with input features averaged over all channels in the bipolar montage at a) 12 hours after CA and b) 24 hours after CA. The blue line represents the mean ROC with the grey area its mean standard deviation. The red dot represents the sensitivity at 100% specificity for prediction of poor outcome. The green dot represents the sensitivity at 95% specificity for prediction of good outcome. The performance at 12 hours after CA shows higher mean values for all performance metrics, but with a wider range for the standard deviation compared to 24 hours after CA.

The baseline performance metrics for the full montage of all electrodes referenced to Cz equal an AUC of 0.912 (0.902-0.923), Se100 of 81% (79-83) for poor outcome and Se95 of 75% (72-77) for prediction of good outcome at 12 hours after CA, and an AUC of 0.910 (0.904-0.916), Se100 of 69% (67-71) for poor outcome and Se95 of 60% (57-62) for good outcome at 24 hours after CA. Corresponding ROC curves are shown in Figure 5, and ROC curves for the reduced montages in Appendix A.6. Eliminating the connectivity features only affects the Se100 at 12 hours after CA negatively. Furthermore, almost all single channels show significantly lower performance than the baseline performance. The performance metrics of the LR model with 95% confidence intervals and their corresponding p-values for the reduced montages compared to the full montage of all electrodes referenced to Cz at 12 hours and 24 hours after CA can be found in Appendix A.7.



a) ROC curve for averaged input over all channels in the refCz montage at 12h after CA



b) ROC curve for averaged input over all channels in the refCz montage at 24h after CA

Figure 5: Mean ROC curve and its standard deviation over 50 iterations with input features averaged over all channels in the refCz montage at a) 12 hours after CA and b) 24 hours after CA. The blue line represents the mean ROC with the grey area its mean standard deviation. The red dot represents the sensitivity at 100% specificity for prediction of poor outcome. The green dot represents the sensitivity at 95% specificity for prediction of good outcome. The performance at 12 hours after CA shows higher mean values for all performance metrics, but with a wider range for the standard deviation compared to 24 hours after CA.

For the frontotemporal montage at 12 hours after CA, the AUC equals 0.902 (0.890-0.913), with a sensitivity of 80% (78-82) at 100% specificity for predicting poor outcome and 73% (70-76) sensitivity at 95% specificity for prediction of good outcome. Sensitivity for prediction of poor outcome is significantly higher in the frontotemporal montage compared to the full bipolar montage. The corresponding ROC curve in comparison to the ROC curve of the full bipolar montage is shown in Figure 6a. Given an AUC of 0.896 (0.889-0.902), Se100 of 68% (67-70) for predicting poor outcome and Se95 of 52% (50-55) for good outcome, the LR model shows better performance on the frontotemporal montage compared to the bipolar montage at 24 hours after cardiac arrest, with a significant increase of AUC and Se100. Corresponding ROC curves are shown in Figure 6b. All values for the performance metrics with 95% confidence intervals and their corresponding p-values for the frontotemporal montage compared to the full bipolar montage are shown in Appendix A.7.

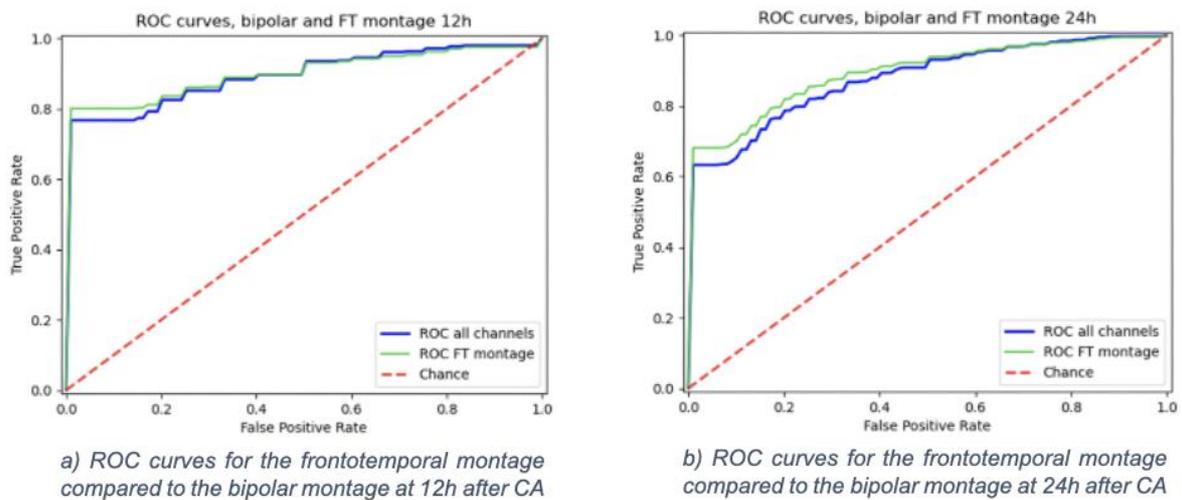


Figure 6: Mean ROC curves over 50 iterations for the frontotemporal montage (green) compared to the full bipolar montage (blue) at a) 12 hours after CA and b) 24 hours after CA. The frontotemporal montage outperforms the bipolar montage at both timepoints.

L1 regularization

Boxplots of all feature values for good and poor outcome were used to determine which features should be modified before summation. All boxplots can be found in Appendix A.8. Four features were inverted before summation for both timepoints: Tsallis entropy, normalized theta power, total signal power and regularity. At 12 hours after CA, the baseline performance without regularizer equals an AUC of 0.871 (0.858-0.883), with Se100 of 73% (70-75) for prediction of poor outcome and Se95 of 65% (61-68) for good outcome. The three most dominant channels found are T3, T4 and C3, with a mean of absolute values of the trained weights of 1.19, 0.91 and 0.67, respectively. Performance metrics at 24 hours after CA without regularizer equal an AUC of 0.837 (0.829-0.845), Se100 of 64% (63-66) for poor outcome and Se95 of 30% (28-33) for good outcome. The channels C3, Fp1 and O2 contributed most to the outcome prediction with mean trained weights of 1.44, 1.20 and 0.99, respectively. For both timepoints, the performance decreases slightly for $\lambda = 0.0005$ compared to no regularization but increases again when the value for λ increases, up to a value of $\lambda = 0.1$. When λ reaches a value of 0.5, the performance decreases again. Remarkably, all performance metrics reach the highest values when all weights, except for one, are in the range of 0.00 to 0.01. Performance metrics and trained weights for all values of λ are shown in Appendix A.9.

4 Discussion

During this study, we aimed to optimize both the monitoring and analysis methods for using cEEG on the ICU as a predictive tool for neurological outcome after CA. The registration quality of the easy-to-apply BrainStatus electrode was not sufficient for accurate scoring. In general, the prognostic value of EEG patterns does not seem to be affected by electrode reduction for both visual assessment and quantitative analysis. The three components of the study, testing the BrainStatus electrode, visual assessment of the EEG based on reduced electrode sets and electrode reduction for qEEG analysis, are discussed separately.

4.1 BrainStatus

In this part of the study, we aimed to evaluate the signal quality and visual assessment of the EEG signal monitored with the BrainStatus electrode, while intermittent check-ups of the electrode placement and impedances were performed. We did not see an explicit dependency between the check-ups and the artefact score of the BrainStatus signals, indicating these check-ups are not affecting the signal quality over time. The classification agreement between the BrainStatus montage and the bipolar montage for the three characteristics background pattern, rhythmic pattern and assigned outcome is moderate ($\kappa = 0.48$), fair ($\kappa = 0.38$) and substantial ($\kappa = 0.61$), respectively. Unfortunately, the study population was very small, partially due to COVID-19 measures and an unusually high number of transfers due to bed shortage. This implies lower reliability of these values.

The classification agreement between the BrainStatus montage and the bipolar montage seems slightly increased compared to earlier visual scoring assessment of BrainStatus EEG epochs without additional check-ups.³¹ The agreement in the previous report equals 0.45 for background pattern, a value of 0.26 for rhythmic pattern and for the assigned outcome prediction the value for kappa was 0.53. The current results imply a slight improvement of the agreement between the BrainStatus montage and the bipolar montage. However, the values for kappa are still moderate, and no significance can be measured due to the minor patient groups and wide range of 95% confidence intervals.

The number of included patients for this part of the study is deficient. Given the small dataset and wide range of the 95% confidence intervals for both ICC and Cohen's kappa, the reliability of these values should be questioned. According to literature, at least 30 heterogeneous samples should be involved to obtain reliable results.⁴⁴ If this requirement is not met, lower values could also reflect the lack of variability among the samples or the small number of subjects. This could also explain the extremely low ICC value for scoring background pattern in the BrainStatus montage (ICC = 0.07), as the patterns are only scored as continuous or not assessable within this montage, which clearly shows a lack of variability. More samples should be included to improve the reliability of ICC and Cohen's kappa. Another consequence of the limited dataset was found in the fact that observers were able to recognize patterns from the same patient in earlier montages. This devaluates the effect of being blinded for patient and patient's condition.

We aimed to obtain a well-distributed variety of background patterns registered with the BrainStatus. It turned out that patients with suppressed or low voltage background patterns were treated with ECMO more often, resulting in exclusion for the BrainStatus study. Therefore, we did not succeed

to include such patterns in this part of the study, resulting in an incomplete representation of the patient population admitted to the ICU after CA.

The 9-channel bipolar montage was used as the golden standard for EEG pattern classification using the visual scoring app in this study. As this montage is used for clinical practice in the AMC, the expectation was to obtain similar classifications from the bipolar montage in the visual scoring app as the patterns described in the patient files. When comparing the results from Figure 3 with patient characteristics in Table 2 the patterns appear to be not in accordance with each other. This shows that the patterns scored in the bipolar montage with the scoring app are not in correspondence with the patterns scored clinically. This might be a result of the limited duration of only five minutes of the EEG epochs used for the visual scoring app, or a result of the preprocessing of the EEG signal.

Based on the results from this study, the suitability for clinical application is not changed compared to previous research with the BrainStatus in the AMC.³⁰⁻³² This means the BrainStatus was still not found to be reliable enough for early prognostication of postanoxic comatose patients at the ICU, also with additional intermittent check-ups. However, the study population is too small to definitely eliminate the BrainStatus as an easy-to-apply EEG possibility.

4.2 Electrode reduction for visual assessment

To evaluate the prognostic value of EEG recordings from a frontotemporal montage, we reassessed EEG recordings from postanoxic comatose patients after cardiac arrest in this reduced montage and compared this to the assessment of EEG patterns in a longitudinal bipolar montage based on 9 electrodes. Reducing the number of electrodes did not affect the interobserver agreement for scoring the background pattern and assigned outcome. Also, the agreement in classification of the EEG background pattern was found to be reliable, and the reduced electrode set showed no significantly different prognostic value for outcome prediction.

Any unfavourable EEG pattern (suppressed, low voltage or burst-suppression with identical bursts) at 24 hours after cardiac arrest was invariably associated with poor neurological outcome. The sensitivity at 100% specificity of 31-34% for predicting poor outcome in this study is slightly lower than some of the earlier reported sensitivities at 100% specificity based on EEG pattern.^{7,10,58,59} However, no difference is seen between the frontotemporal montage and the bipolar montage within this study, suggesting this lower sensitivity is not affected by the reduced montage. We found a continuous EEG pattern with normal voltage at 12 hours after CA to be strongly related with good neurological outcome in both the bipolar and the frontotemporal montage with sensitivities of 64% and 59%, respectively. These values are in accordance with earlier reported outcome prediction studies.^{7,10,58}

Several studies explored the quality of EEG monitoring with a reduced number of electrodes before. Tjepkema-Cloostermans et al.⁵⁸ showed a successful reduction from 19 to 10 electrodes, without a decrease of prognostic value for neurological outcome after CA. Also, amplitude-integrated EEG based on two bipolar channels from 4 central electrodes was found to contain relevant information for outcome prediction in patients with postanoxic coma.^{60,61} Some studies specifically explored the possibilities for reducing to sub-hairline montages. Kortelainen et al.³⁴ reported a reduced set with electrodes at locations Fp1, Fp2, F7 and F8 to be able to capture propofol-induced slow-wave activity

with comparable performance as a full 19-channel EEG cap. These results support the agreement found between the bipolar and the frontotemporal montage in this study. Additionally, Tanner et al.³⁶ investigated a sub-hairline electrode set with 6 EEG electrodes for classification of EEG recordings according to the Young-McLachlan classification system and for seizure detection. In general, the agreement between the two montages was only fair with a kappa of 0.30 (95% CI: 0.21-0.40). However, they state that sub-hairline EEG might be able to meet the clinical requirements in patients with less heterogeneous background and well-defined clinical questioning, which is the case in patients with postanoxic encephalopathy.

The frontotemporal montage proved to be useful for outcome prediction of postanoxic comatose patients based on classification of EEG background patterns. From clinical perspective, it is essential that the absence of false positives for the prediction of poor outcome is still achieved within the reduced montage. It should be kept in mind that electrode reduction up to the usage of only four locations also increases the risk of not obtaining valid registrations. If only one or two electrodes fail to register proper signal due to artefacts or loosening electrodes, the classification of the EEG pattern will immediately be impeded.

In this part of the study, we focused on the visual assessment of the EEG pattern based on a reduced montage. The prognostic value of the EEG background pattern is evaluated isolated from other clinical characteristics. To obtain reliable prognosis and diagnostic value, the EEG pattern should always be interpreted within a multimodal approach, also accounting for other clinical characteristics, such as somatosensory evoked potentials.^{2,10,41,46}

In conclusion, we can state that the prognostic value of EEG background patterns in postanoxic comatose patients at the ICU is not affected by reducing the number of electrodes from a 9-channel bipolar montage to a 4-channel frontotemporal montage. Using 4 electrodes at sub-hairline locations creates the opportunity to further explore the possibilities of using easy-to-apply adhesive electrodes for cEEG monitoring in the ICU.

4.3 Electrode reduction for qEEG analysis

In this part of the study, we evaluated the performance of a logistic regression model for outcome prediction on several reduced electrode sets, based on AUC of the ROC, Se100 for prediction of poor outcome and Se95 for good outcome prediction. Reduction from 8 to 2 channels within the bipolar montage did not decrease the model's performance for prediction at 12 hours after CA. At 24 hours after CA, only two of the reduced montages showed decreased performance, and two other reduced montages showed increased performance metrics. The full refCz montage showed better performance than the bipolar montage. However, single channels extracted from this montage almost all showed less predictive value for neurological outcome. The frontotemporal montage showed a higher sensitivity for poor outcome prediction for both timepoints without a decrease of other performance metrics.

By applying L1 regularization while using the sum of all features for single channels as input for the model, we show that the contribution to the prediction is not equal for all channels. At 12 hours after CA, the three most dominant channels found are T3, T4 and C3. Comparing this to the performance of the reduced montages, we see that the channels adjacent to T4 in the bipolar montage show increased

performance. At 24 hours after CA, L1 regularization shows that C3 contributes the most to the outcome prediction, and according to that, the channels adjacent to C3 show increased performance. On the contrary, the channels adjacent to O2 show decreased performance, while this channel also shows a high contribution to the outcome prediction according to the value of the weights. Furthermore, Fp1 showed a high contribution to the model's prediction at 24 hours after CA without changed performance for the reduced montage of its adjacent channels. Based on this inconsistency between performance on reduced montages and the weights per channel in L1 regularization, it remains unclear which electrode locations contribute most to the outcome prediction model. However, 3 of the 4 electrode locations from the frontotemporal montage (Fp1, T3 and T4) are represented in the top three most contributing channels at 12 and 24 hours after CA. In combination with the increased model's performance on this input, this speaks in favour of the frontotemporal montage as useful electrode locations for outcome prediction using a LR model.

In some cases, the predictive value of the model increases when reduced montages are used as the input. For all montages containing multiple channels, the features were averaged over all used channels, raising the question if information is lost by averaging over too many channels. On the contrary, the model's performance decreases for almost all single channel inputs, indicating a lack of information for accurate outcome prediction. Further investigation is needed to find the optimal number of electrodes for this model.

The baseline LR model, with input from all channels in a bipolar montage, outperformed most of the earlier developed outcome prediction algorithms.^{8,20,21,23,24,38,62} Nagaraj et al.²⁰ reported the best overall performance found in literature with an AUC of 0.94, Se100 of 0.66 for poor outcome and Se95 of 0.72 for good outcome at 12 hours after CA. This outperforms our model on the AUC for all montages and on Se95 for good outcome when using the bipolar montage as input. At 24 hours after CA, Nagaraj et al. reported an AUC of 0.88, Se100 of 0.60 and Se95 of 0.40, which are all lower than the performance metrics found in this study. Detailed comparison of the performance metrics found in literature can be found in Appendix A.10.

The model used in this study was based on the LR model from an earlier study in the AMC by Van Poppel²⁴ and showed comparable results, except for the prediction of good outcome at 12 hours after CA (0.30 vs 0.71). This can be assigned to a calculation error and confusion between the false positive rate and the specificity in Van Poppel's research, which we restored in this study. From clinical perspective, it makes sense that the sensitivity of predicting good outcome declines over time. Several papers reported the evolution of EEG background pattern over time, including Oh et al.¹⁶ They showed that almost no patients with poor neurological outcome had a continuous EEG background pattern at 12 hours after CA, and from 20 hours after CA, this number of patients is increasing, resulting in a continuous background pattern becoming less specific for good neurological outcome.

Based on the performance of the LR model on reduced electrode sets from the bipolar montage, the assumption can be made that just a few electrodes can gather enough information for a functional outcome prediction. This supports the potential of using a reduced (frontotemporal) montage in clinical practice. The prognostic values at 12 hours after CA are higher than at 24 hours after CA for all performance metrics and all montages. This emphasizes the importance of early start of cEEG

registrations in the ICU. The suitability of reduced montages for outcome prediction creates opportunities for the use of easy-to-apply adhesive electrodes for cEEG monitoring in the ICU, which will contribute to earlier start of monitoring.

The model's performance based on reduced montages of only two channels from the bipolar montage is promising. However, the model was trained and tested on a limited dataset, especially for 12 hours after CA, and this data was obtained from only one centre. External validation with larger datasets is required to validate the LR model with full montages and reduced montages further.

5 Recommendations

Results from this study show high potential for the use of sub-hairline EEG montages for prognostication after CA for both visual scoring and qEEG analysis. Yet, the BrainStatus electrode set does not fulfil the requirements for adequate EEG monitoring at this point. To find alternative easy-to-apply EEG set-ups, future research could focus on using alternative, adhesive electrodes from other domains like ECG and EMG to register EEG signals with sufficient signal quality. Preliminary research into several adhesive electrodes showed promising results for the Ambu® Neuroline 720 regarding impedances and signal quality.³³ These electrodes were only tested on healthy subjects for a short period of time, and no information is known about the suitability in an ICU setting yet. We would recommend continuing this research and start monitoring postanoxic comatose patients in the ICU with these electrodes, simultaneously with the conventional cup electrodes.

We showed that the prognostic value of EEG background patterns in postanoxic comatose patients at the ICU is not affected by reducing the number of electrodes from a 9-channel bipolar montage to a 4-channel frontotemporal montage. This suggests that clinical application of reduced montages might be possible in the future. A possible way to start implementing this could be to start EEG monitoring as soon as possible with the reduced montage. If any irregularities are seen on the reduced EEG or the signal cannot be appropriately assessed, the EEG montage can be expanded to the full bipolar montage.

The performance of the LR model on reduced montages seems very promising. To confirm the validity of these results, an external validation should be performed. We would also recommend a multi-centre approach to test the model on EEG registrations from different centres. Expansion of the dataset is recommended to obtain more robust results, especially for the dataset at 12 hours after CA.

6 Conclusion

We showed that intermittent check-ups did not improve the signal quality of the BrainStatus, which is required to use the BrainStatus as a reliable monitoring tool. However, the prognostic value of visual scoring of EEG patterns is unchanged by reducing the number of electrodes from 9 to 4. Also, the qEEG analysis showed significant better performance for the 4-channel frontotemporal montage compared to the conventional 9-channel bipolar montage. This strongly suggests a high suitability of this montage for clinical use in prognostication for postanoxic comatose patients. Overall, the performance of the LR model was preserved while using reduced montages of only two channels as input.

References

1. Bernard SA, Gray TW, Buist MD, et al. Treatment of comatose survivors of out-of-hospital cardiac arrest with induced hypothermia. *N Engl J Med.* 2002;346(8):557-563.
2. Oddo M, Rossetti AO. Early multimodal outcome prediction after cardiac arrest in patients treated with hypothermia. *Crit Care Med.* 2014;42(6):1340-1347. doi:10.1097/CCM.0000000000000211
3. Nielsen N, Wetterslev J, Cronberg T, et al. Targeted temperature management at 33 °C versus 36 °C after cardiac arrest. *N Engl J Med.* 2013;369(23):2197-2206. doi:10.1056/NEJMoa1310519
4. 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(1):1-8. doi:10.1186/s13054-017-1693-2
5. Ruijter BJ, Tjepkema-Cloostermans MC, Tromp SC, et al. Early electroencephalography for outcome prediction of postanoxic coma: A prospective cohort study. *Ann Neurol.* 2019;86(2):203-214. doi:10.1002/ana.25518
6. Cloostermans MC, van Meulen FB, Eertman CJ, Hom HW, van Putten MJAM. Continuous electroencephalography monitoring for early prediction of neurological outcome in postanoxic patients after cardiac arrest: A prospective cohort study. *Crit Care Med.* 2012;40(10):2867-2875. doi:10.1097/CCM.0b013e31825b94f0
7. Muhlhofer W, Szaflarski J. Prognostic Value of EEG in Patients after Cardiac Arrest—An Updated Review. *Curr Neurol Neurosci Rep.* 2018;18(4). doi:10.1007/s11910-018-0826-6
8. Tjepkema-Cloostermans M, Hofmeijer J, Beishuizen A, et al. Cerebral recovery index: Reliable help for prediction of neurologic outcome after cardiac arrest. *Crit Care Med.* 2017;45(8):789-797. doi:10.1097/CCM.0000000000002412
9. Tjepkema-Cloostermans M, van Meulen F, Meinsma G, van Putten M. A Cerebral Recovery Index (CRI) for early prognosis in patients after cardiac arrest. *Crit Care.* 2013;17(5):1. doi:10.1186/cc13078
10. Hofmeijer J, Beernink T, Bosch F, Beishuizen A, Tjepkema-Cloostermans M, Van Putten M. Early EEG contributes to multimodal outcome prediction of postanoxic coma. *Neurology.* 2015;85(2):137-143. doi:10.1212/WNL.0000000000001742
11. Citerio G, Patruno A, Beretta S, Longhi L, Frigeni B, Lorini L. Implementation of continuous qEEG in two neurointensive care units by intensivists: a feasibility study. *Intensive Care Med.* 2017;43(7):1067-1068. doi:10.1007/s00134-017-4775-3
12. Rijdsdijk M, Leijten F, Slooter A. Continuous EEG monitoring in the Intensive Care Unit. *Netherlands J Crit Care.* 2008;12(4):157-162. doi:10.1016/B978-0-444-63600-3.00007-6
13. Hilkmann D, Van Mook W, Van Kranen-Mastenbroek V. Continuous electroencephalographic-monitoring in the ICU: An overview of current strengths and future challenges. *Curr Opin Anaesthesiol.* 2017;30(2):192-199. doi:10.1097/ACO.0000000000000443
14. Hilkmann D, van Mook W, Mess W, van Kranen-Mastenbroek V. The Use of Continuous EEG Monitoring in Intensive Care Units in The Netherlands: A National Survey. *Neurocrit Care.* 2018;29(2):195-202. doi:10.1007/s12028-018-0525-9
15. Foreman B, Claassen J. Quantitative EEG for the detection of brain ischemia. *Crit Care.* 2012;16(2). doi:10.1186/cc11230
16. Oh SH, Park KN, Shon YM, et al. Continuous Amplitude-Integrated Electroencephalographic Monitoring Is a Useful Prognostic Tool for Hypothermia-Treated Cardiac Arrest Patients. *Circulation.* 2015;132(12):1094-1103. doi:10.1161/CIRCULATIONAHA.115.015754
17. Westhall E, Rosén I, Rossetti AO, et al. Interrater variability of EEG interpretation in comatose cardiac arrest patients. *Clin Neurophysiol.* 2015;126(12):2397-2404. doi:10.1016/j.clinph.2015.03.017
18. Benarous L, Gavaret M, Soda Diop M, et al. Sources of interrater variability and prognostic value of standardized EEG features in post-anoxic coma after resuscitated cardiac arrest. *Clin Neurophysiol Pract.* 2019;4:20-26. doi:10.1016/j.cnp.2018.12.001
19. Scheuer ML, Wilson SB. Data analysis for continuous EEG monitoring in the ICU: Seeing the forest and the trees. *J Clin Neurophysiol.* 2004;21(5):353-378. doi:10.1097/01.WNP.0000147078.82085.D3
20. Nagaraj SB, Tjepkema-Cloostermans MC, Ruijter BJ, Hofmeijer J, van Putten MJAM. The revised Cerebral Recovery Index improves predictions of neurological outcome after cardiac arrest. *Clin Neurophysiol.* 2018;129(12):2557-2566. doi:10.1016/j.clinph.2018.10.004
21. Jonas S, Rossetti AO, Oddo M, Jenni S, Favaro P, Zubler F. EEG-based outcome prediction after cardiac arrest with convolutional neural networks: Performance and visualization of discriminative features. *Hum Brain Mapp.* 2019;40(16):4606-4617. doi:10.1002/hbm.24724
22. Craik A, He Y, Contreras-Vidal JL. Deep learning for electroencephalogram (EEG) classification tasks: A review. *J Neural Eng.* 2019;16(3). doi:10.1088/1741-2552/ab0ab5
23. Tjepkema-Cloostermans MC, da Silva Lourenço C, Ruijter BJ, et al. Outcome Prediction in Postanoxic Coma With Deep Learning. *Crit Care Med.* 2019;47(10):1424-1432. doi:10.1097/CCM.0000000000003854
24. van Poppel LM. EEG-Based Neurological Outcome Prediction after Cardiac Arrest with an LSTM RNN: Can We Predict Life after the Heart Stops Beating? *Master Thesis, Delft Univ Technol.* 2020.
25. Rubin MN, Jeffery OJ, Britton JW, et al. Efficacy of a Reduced Electroencephalography Electrode Array for Detection of Seizures. *The Neurohospitalist.* 2014;4(1):6-8. doi:10.1177/1941874413507930
26. Lepola P. Novel EEG Electrode Set for Emergency Use. *PhD thesis, East Univ Finl.* 2014.

27. University of Eastern Finland. New EEG electrode set for fast, easy measurement of brain function abnormalities. *ScienceDaily*. 2014;20-22. www.sciencedaily.com/releases/2014/09/140924084847.htm.
28. Lepola P, Myllymaa S, Töyräs J, et al. Screen-printed EEG electrode set for emergency use. *Sensors Actuators, A Phys*. 2014;213:19-26. doi:10.1016/j.sna.2014.03.029
29. Myllymaa S, Lepola P, Töyräs J, et al. New disposable forehead electrode set with excellent signal quality and imaging compatibility. *J Neurosci Methods*. 2013;215(1):103-109. doi:10.1016/j.jneumeth.2013.02.003
30. Lievestro H. Bedside quantitative cEEG monitoring on the Intensive Care for comatose patients after cardiac arrest. *Master Thesis, Univ Twente*. 2019.
31. van Merkerk MN. Assessment of the suitability of using a forehead electrode set for continuous EEG monitoring in comatose patients after cardiac arrest. *Rep Clin Internship, Univ Twente*. 2020.
32. van Heijst R. Characterization of artefacts present in a forehead electrode set for continuous EEG monitoring in comatose patients after cardiac arrest and the development of an alarmsystem to detect poor signal quality. *Rep Clin Internship, Univ Twente*. 2020.
33. Edgar R. Towards New Methods of Continuous EEG Monitoring in Comatose Patients after Cardiac Arrest at the Intensive Care. *Rep Clin Internship, Univ Twente*. 2020.
34. Kortelainen J, Väyrynen E, Juuso I, Laurila J, Koskenkari J, Ala-Kokko T. Forehead electrodes sufficiently detect propofol-induced slow waves for the assessment of brain function after cardiac arrest. *J Clin Monit Comput*. 2020;34(1):105-110. doi:10.1007/s10877-019-00282-3
35. Backman S, Cronberg T, Rosén I, Westhall E. Reduced EEG montage has a high accuracy in the post cardiac arrest setting. *Clin Neurophysiol*. 2020;131(9):2216-2223. doi:10.1016/j.clinph.2020.06.021
36. Tanner AEJ, Särkelä MOK, Virtanen J, et al. Application of subhairline EEG montage in intensive care unit: Comparison with full montage. *J Clin Neurophysiol*. 2014;31(3):181-186. doi:10.1097/WNP.0000000000000049
37. Young GB, Sharpe MD, Savard M, Al Thenayan E, Norton L, Davies-Schinkel C. Seizure detection with a commercially available bedside EEG monitor and the subhairline montage. *Neurocrit Care*. 2009;11(3):411-416. doi:10.1007/s12028-009-9248-2
38. Kortleve M. Reliable prediction of neurological outcome in postanoxic coma patients using EEG and machine learning with only three electrodes. *Rep Res Proj Univ Amsterdam*. 2020.
39. Admiraal MM. EEG reactivity for prognostication after cardiac arrest. *PhD thesis, Univ van Amsterdam*. 2019:3-16.
40. Sandroni C, D'Arrigo S, Nolan JP. Prognostication after cardiac arrest. *Crit Care*. 2018;22(1):1-9. doi:10.1186/s13054-018-2060-7
41. Prognose van postanoxisch coma - Richtlijnen database. Date Accessed: 2020-03-13. https://richtlijnen database.nl/richtlijn/prognose_van_postanoxisch_coma/samenvatting_overkoepelende_a_dviezen_bij_de_richtlijn.html. Accessed March 13, 2020.
42. Sessler CN, Gosnell MS, Grap MJ, et al. The Richmond Agitation-Sedation Scale: Validity and reliability in adult intensive care unit patients. *Am J Respir Crit Care Med*. 2002;166(10):1338-1344. doi:10.1164/rccm.2107138
43. Hirsch L, Laroche S, Gaspard N, et al. American clinical neurophysiology society's standardized critical care EEG terminology: 2012 version. *J Clin Neurophysiol*. 2013;30(1):1-27. doi:10.1097/WNP.0b013e3182784729
44. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012
45. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics*. 1977;33(1):159-174. doi:10.2307/2529310
46. Admiraal MM, van Rootselaar AF, Hofmeijer J, et al. Electroencephalographic reactivity as predictor of neurological outcome in postanoxic coma: A multicenter prospective cohort study. *Ann Neurol*. 2019;86(1):17-27. doi:10.1002/ana.25507
47. Ghassemi MM. Life After Death: Techniques for the Prognostication of Coma Outcomes after Cardiac Arrest. 2018:1-134.
48. Thakor N V, Tong S. Advances in Quantitative Electroencephalogram Analysis Methods. *Annu Rev Biomed Eng*. 2004;6(1):453-495. doi:10.1146/annurev.bioeng.5.040202.121601
49. Tsallis C. Possible generalization of Boltzmann-Gibbs statistics. *J Stat Phys*. 1988;52:479-487.
50. Hegger R, Kantz H. Improved false nearest neighbor method to detect determinism in time series data. *Phys Rev E - Stat Physics, Plasmas, Fluids, Relat Interdiscip Top*. 1999;60(4 B):4970-4973. doi:10.1103/physreve.60.4970
51. Geocadin RG, Muthuswamy J, Sherman DL, Thakor N V., Hanley DF. Early electrophysiological and histologic changes after global cerebral ischemia in rats. *Mov Disord*. 2000;15(SUPPL. 1):14-21. doi:10.1002/mds.870150704
52. Box J, Jenkins G, Reinsel G, Ljung G. *Autoregressive Processes*. 5th ed. John Wiley & Sons; 2016.
53. Mack S, Kandel E, Schwartz J, Jessel T, Siegelbaum S, Hudspeth A. *Principles of Neural Science*. 5th ed. McGraw-Hill; 2013.
54. Marieb E, Hoehn K. *Human Anatomy & Physiology*. 9th ed. Pearson; 2019.
55. Nunez P, Srinivasan R. *Electric Fields of the Brain: The Neurophysics of EEG*. 2nd ed. Oxford University Press; 2006.
56. Ghassemi MM, Amorim E, Alhanai T, et al. Quantitative Electroencephalogram Trends Predict Recovery in Hypoxic-Ischemic Encephalopathy. *Crit Care Med*. 2019;47(10):1416-1423.

- doi:10.1097/CCM.0000000000003840
57. Stam CJ, Nolte G, Daffertshofer A. Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Hum Brain Mapp.* 2007;28(11):1178-1193. doi:10.1002/hbm.20346
 58. Tjepkema-Cloostermans MC, Hofmeijer J, Hom HW, Bosch FH, Van Putten MJAM. Predicting outcome in postanoxic coma: Are ten EEG electrodes enough? *J Clin Neurophysiol.* 2017;34(3):207-212. doi:10.1097/WNP.0000000000000337
 59. 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(7):1264-1272. doi:10.1007/s00134-015-3834-x
 60. Rundgren M, Westhall E, Cronberg T, Rosén I, Friberg H. Continuous amplitude-integrated electroencephalogram predicts outcome in hypothermia-treated cardiac arrest patients. *Crit Care Med.* 2010;38(9):1838-1844. doi:10.1097/CCM.0b013e3181eaa1e7
 61. Friberg H, Westhall E, Rosén I, Rundgren M, Nielsen N, Cronberg T. Clinical review: Continuous and simplified electroencephalography to monitor brain recovery after cardiac arrest. *Crit Care.* 2013;17(4). doi:10.1186/cc12699
 62. Derksen M. Quantitative EEG Features to Predict Neurological Outcome in Post-anoxic Coma after Cardiac Arrest: An external validation study. *Rep Res Proj Univ Amsterdam.* 2020.

Appendices

Appendix A.1 BrainStatus

The BrainStatus is a disposable screen-printed electrode set that is easy to apply thus suitable for emergency cases. It consists of ten EEG electrodes, two electro-oculography (EOG) electrodes, two ground electrodes and two reference electrodes. All electrodes were coated with an adhesive hydrogel membrane to improve contact with the skin. Figure A.1 shows an overview of the BrainStatus.

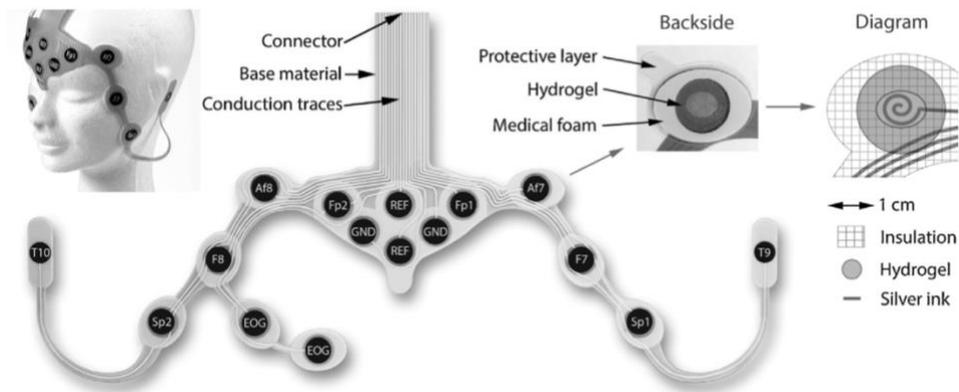


Figure A.1: An overview of the BrainStatus, the electrode placement on a subject's head (left), a backside view and diagram of a single electrode (right). Image adapted from Lepola et al.²⁸

Appendix A.2. User interface scoring application

Six different characteristics were assessed for all EEG samples with the scoring application: rhythmic pattern, rhythmic abundance, background pattern, background voltage, background frequency and whether the assessment of the signal was obscured by the presence of artefacts. Rhythmic pattern could be scored as a spike-wave pattern, periodic discharges, rhythmic delta or none. If any of the patterns were present, their abundance was scored as well. Independently, the background pattern was scored as suppressed (all activity below 10 μV), burst-suppression with or without identical bursts (50-99% suppressed), discontinuous (10-49% suppressed) or continuous. In case of a continuous background pattern, a distinction was made between low voltage (<20 μV) or normal voltage (>20 μV), and the background frequency had to be scored. To indicate the presence of artefacts obscuring the assessment, a check box could be ticked. The user interface with an example EEG is shown in Figure A.2.

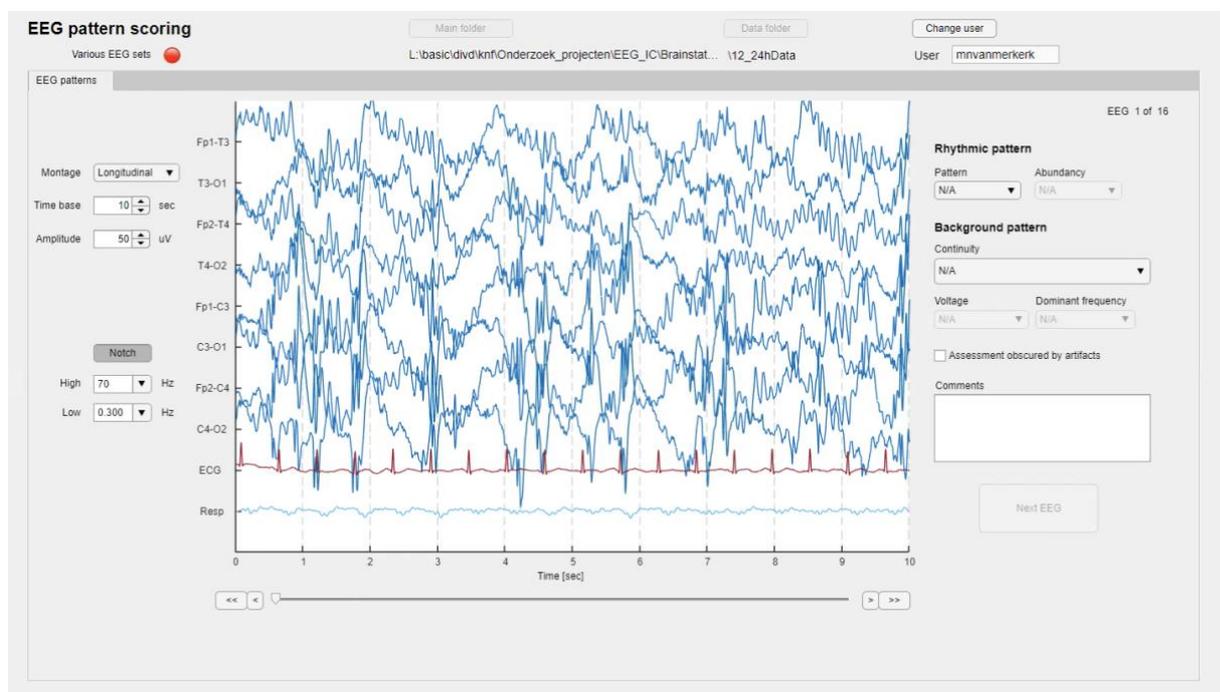


Figure A.2: User interface of the web application for visual scoring of EEG, with example EEG.

Appendix A.3. Additional results BrainStatus

Inter-rater reliability

The inter-rater reliability and its 95% confidence interval for the three main characteristics and all montages are shown in Table A.1.

Table A.1: ICC values representing inter-rater reliability and strength of agreement between observers.

Montage	Characteristic	ICC (95% CI)	Agreement
9-channel bipolar	Background pattern	0.82 (0.37-0.97)	Good
	Rhythmic pattern	0.80 (0.27-0.96)	Good
	Outcome	0.71 (-0.10-0.95)	Moderate
4-channel frontotemporal	Background pattern	0.87 (0.56-0.98)	Good
	Rhythmic pattern	0.79 (0.29-0.96)	Good
	Outcome	0.97 (0.90-0.99)	Excellent
BrainStatus	Background pattern	0.07 (-3.24-0.84)	Poor
	Rhythmic pattern	0.71 (-0.21-0.95)	Moderate
	Outcome	0.75 (0.12-0.95)	Good

Confusion matrices

The confusion matrices for the three main characteristics (background pattern, rhythmic pattern and outcome) in all montage combinations (BrainStatus – bipolar, BrainStatus – frontotemporal and frontotemporal – bipolar) are shown in Figures A.3-A.11. For all figures, N/A is not assessable, blue colour indicates conformity, red represents disagreement and colour saturation corresponds with the relative number of observations.

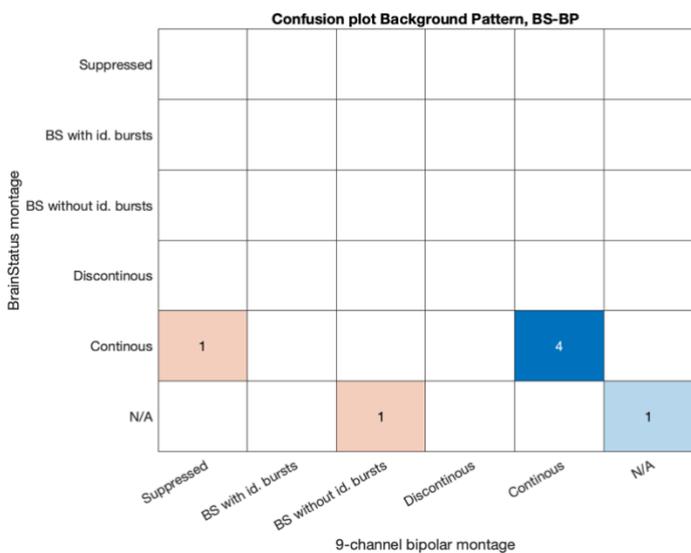


Figure A.3: Confusion matrix for scored EEG background patterns comparing the BrainStatus montage with the bipolar montage

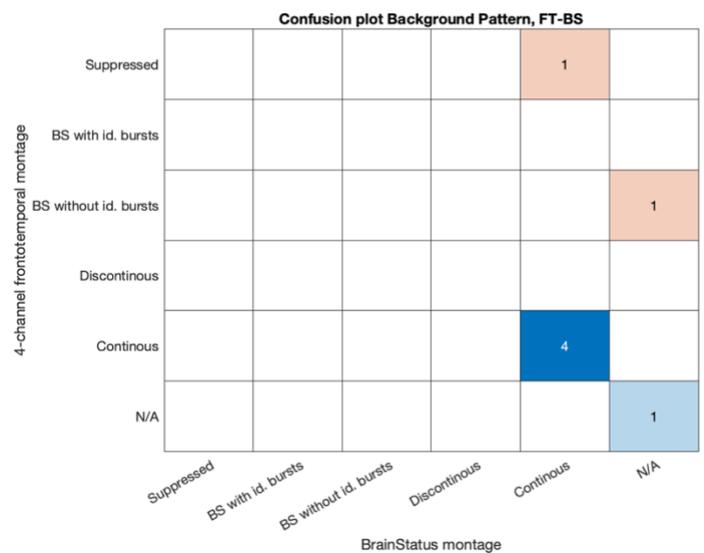


Figure A.4: Confusion matrix for scored EEG background patterns comparing the frontotemporal montage with the BrainStatus montage

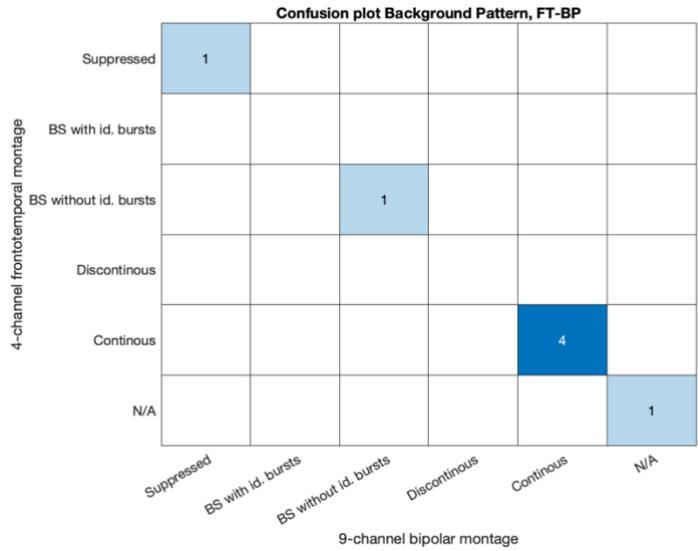


Figure A.5: Confusion matrix for scored EEG background patterns comparing the frontotemporal montage with the bipolar montage

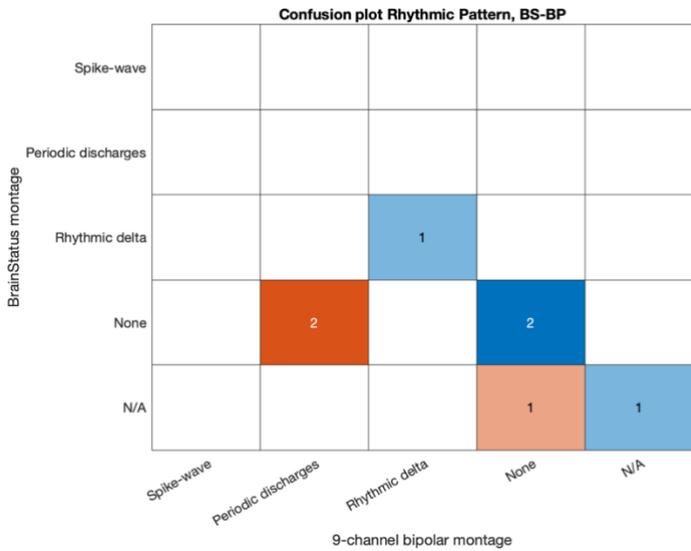


Figure A.6: Confusion matrix for scored EEG rhythmic patterns comparing the BrainStatus montage with the bipolar montage

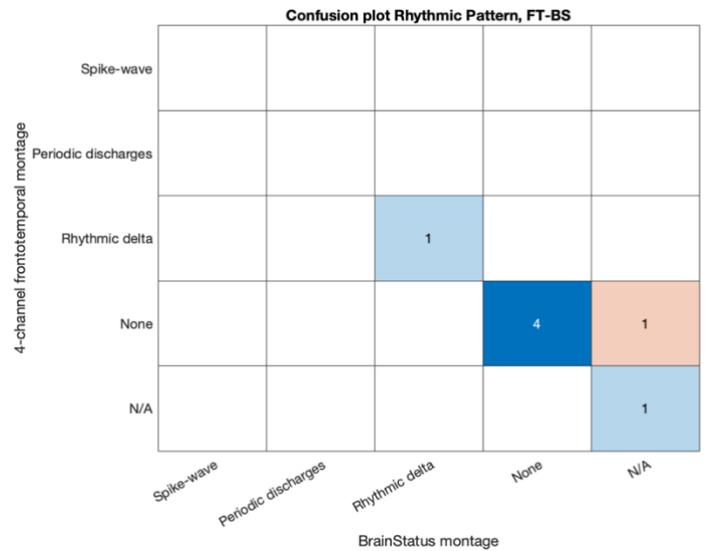


Figure A.7: Confusion matrix for scored EEG rhythmic patterns comparing the frontotemporal montage with the BrainStatus montage

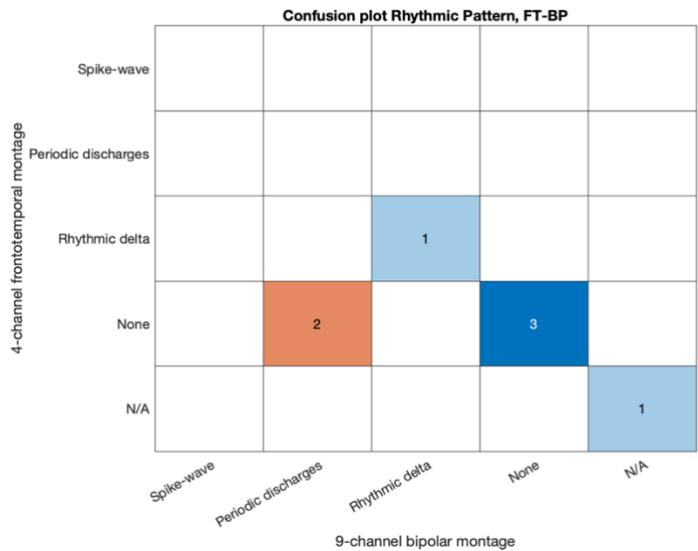


Figure A.8: Confusion matrix for scored EEG rhythmic patterns comparing the frontotemporal montage with the bipolar montage

Confusion plot Outcome, BS-BP

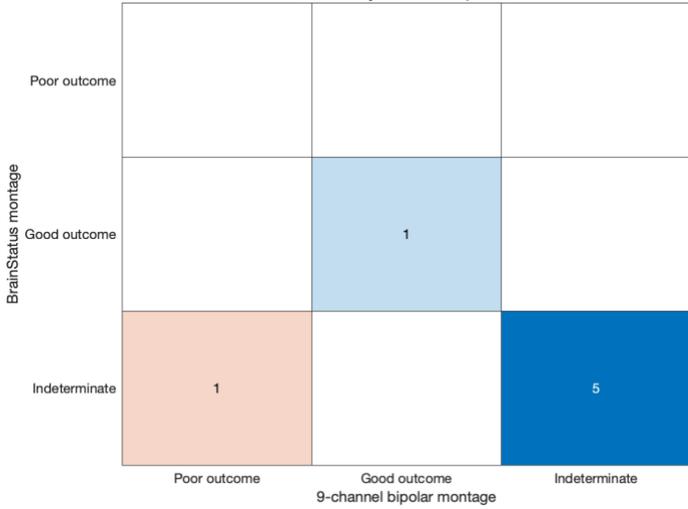


Figure A.10: Confusion matrix for assigned outcome prediction comparing the BrainStatus montage with the bipolar montage

Confusion plot Outcome, FT-BS

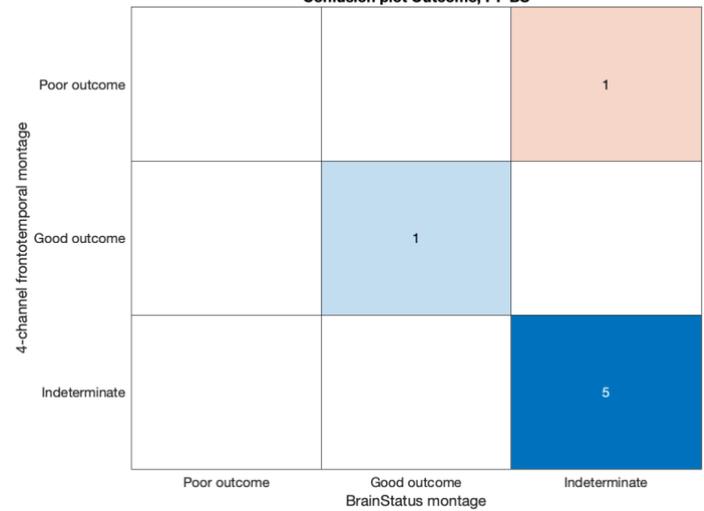


Figure A.9: Confusion matrix for assigned outcome prediction comparing the frontotemporal montage with the BrainStatus montage

Confusion plot Outcome, FT-BP

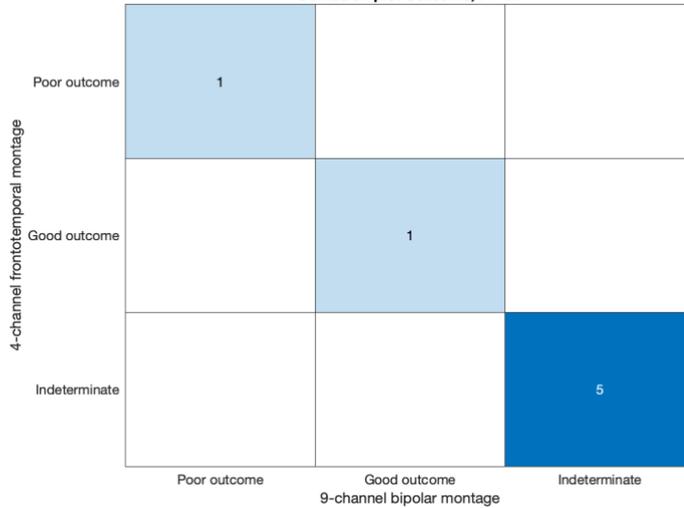


Figure A.11: Confusion matrix for assigned outcome prediction comparing the frontotemporal montage with the bipolar montage

Classification agreement

The calculated values for Cohen's kappa for background pattern, rhythmic pattern and assigned outcome for all montage combinations are shown in Table A.2.

Table A.2: Classification agreement between the three electrode sets

Montages	Characteristic	Cohen's κ (95% CI)	Agreement
BrainStatus - bipolar	Background pattern	0.48 (-0.13-1.09)	Moderate
	Rhythmic pattern	0.38 (-0.15-0.91)	Fair
	Outcome	0.61 (-0.09-1.32)	Substantial
Frontotemporal - BrainStatus	Background pattern	0.48 (-0.13-1.09)	Moderate
	Rhythmic pattern	0.73 (0.24-1.22)	Substantial
	Outcome	0.61 (-0.09-1.32)	Substantial
Frontotemporal - bipolar	Background pattern	1.00 (1.00-1.00)	Perfect
	Rhythmic pattern	0.57 (0.05-1.07)	Moderate
	Outcome	1.00 (1.00-1.00)	Perfect

Appendix A.4. Artefact scores BrainStatus

Figure A.12 shows the artefact score over time plotted for all patients separately. For patient 2 and patient 4 (Figure A.12b and A.12d), the artefact score for the Brainstatus increases clearly over time and increases more rapidly than the conventional cup electrodes. For patient 3, the monitoring with the BrainStatus was started later than the conventional monitoring, indicated in Figure A.12c.

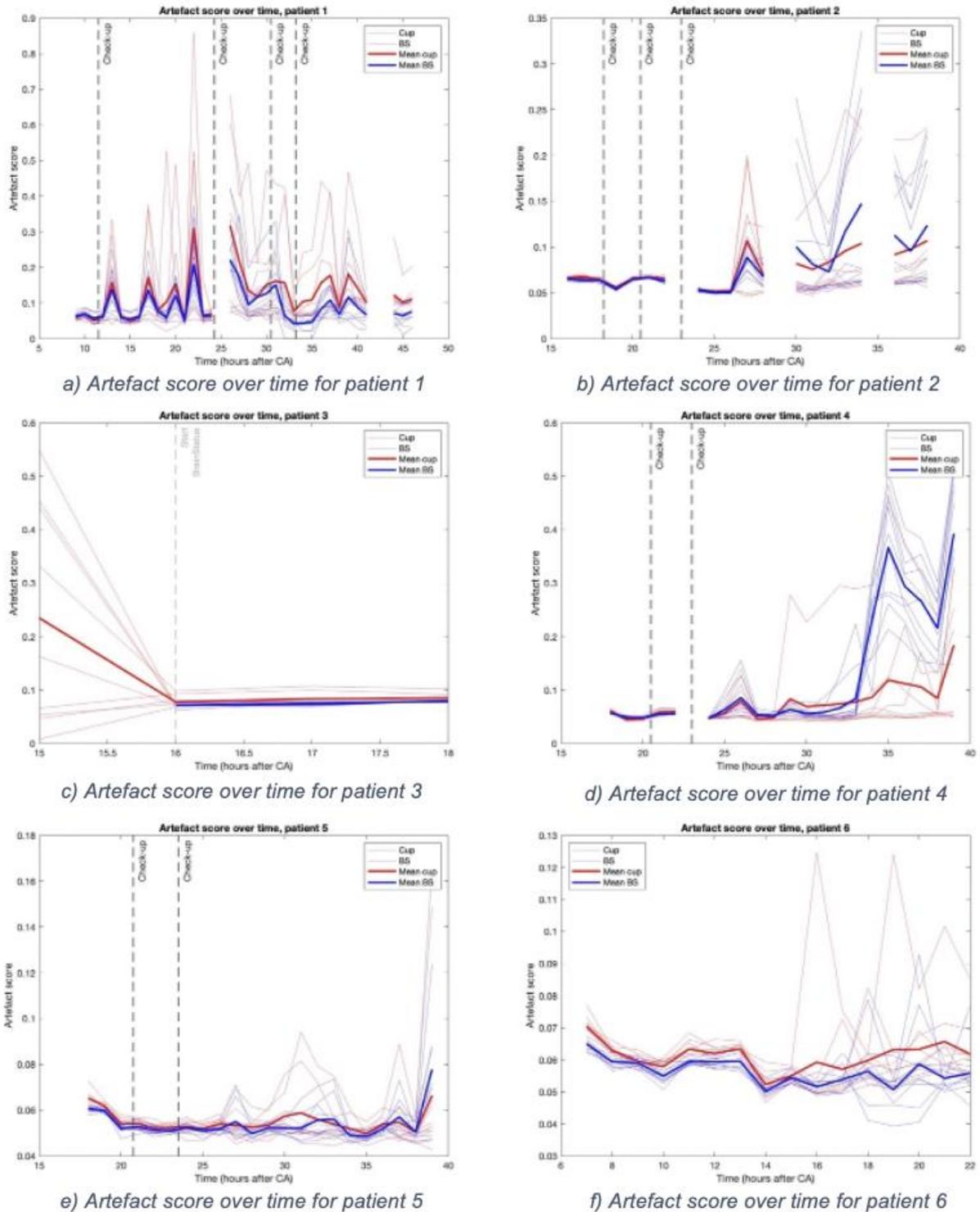


Figure A.12: Plots for the calculated artefact score over time for patient 1-6 in respectively Figures A.12a-A.12f. The lighter red lines represent single channels registered with the conventional cup electrodes, the lighter blue lines represent single channels from the BrainStatus electrode. The thicker red and blue lines represent the mean value of all single channels from the cup electrodes and the BrainStatus electrodes respectively. Black dashed lines indicate the check-ups of the BrainStatus. The artefact score is not consistently higher for the BrainStatus compared to the cup electrodes. In only two patients, the artefact score is clearly increasing over time for both electrodes.

Appendix A.5. Additional results electrode reduction for visual assessment

Inter-rater reliability

The inter-rater reliability and its 95% confidence interval for the three main characteristics and both montages are shown in Table A.3.

Table A.3: ICC values representing inter-rater reliability and strength of agreement between observers.

Montage	Characteristic	ICC (95% CI)	Agreement
9-channel bipolar	Background pattern	0.94 (0.93 – 0.95)	Excellent
	Rhythmic pattern	0.81 (0.77 – 0.85)	Good
	Outcome	0.91 (0.89 – 0.93)	Excellent
4-channel frontotemporal	Background pattern	0.94 (0.92 – 0.95)	Excellent
	Rhythmic pattern	0.71 (0.63 – 0.77)	Moderate
	Outcome	0.89 (0.87 – 0.92)	Good

ICC: intra-class correlation coefficient; CI: confidence interval

Confusion matrices

Additional confusion matrices for rhythmic pattern and outcome are shown in Figures A.13 and A.14. For all figures, N/A is not assessable, blue colour indicates conformity, red represents disagreement and colour saturation corresponds with the relative number of observations.

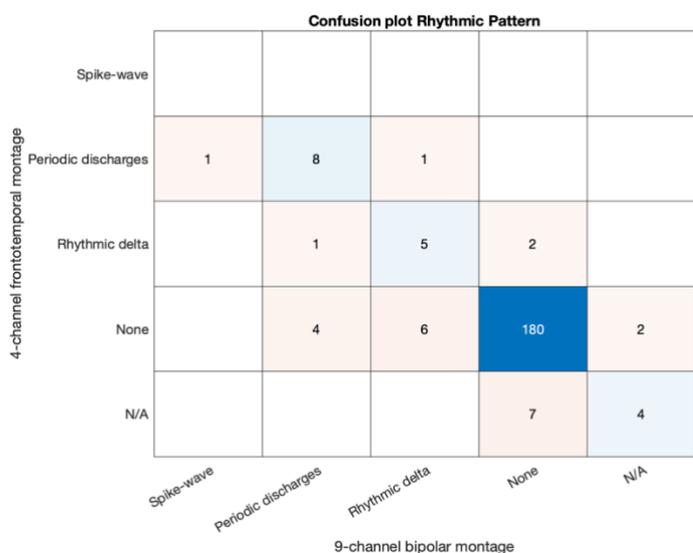


Figure A.13: Confusion matrix for scored EEG rhythmic patterns comparing the frontotemporal montage with the bipolar montage



Figure A.14: Confusion matrix for assigned outcome prediction comparing the frontotemporal montage with the bipolar montage

Classification agreement

The calculated values for Cohen's kappa for background pattern, rhythmic pattern and assigned outcome for all montage combinations are shown in Table A.4.

Table A.4: Classification agreement between the frontotemporal montage and the bipolar montage.

Montages	Characteristic	Cohen's κ (95% CI)	Agreement
Frontotemporal - bipolar	Background pattern	0.76 (0.69 – 0.83)	Substantial
	Rhythmic pattern	0.57 (0.40 – 0.73)	Moderate
	Outcome	0.86 (0.78 – 0.94)	Almost perfect

Appendix A.6. ROC curves reduced montages

ROC curves for all reduced montages are shown in Figure A.15. The reduced montages compared to the full bipolar montage are shown in Figure A.15a and A.15b for 12 and 24 hours after CA, respectively. The single channel montages compared to all channels referenced to Cz are shown in Figure A.15c and A.15d for both timepoints.

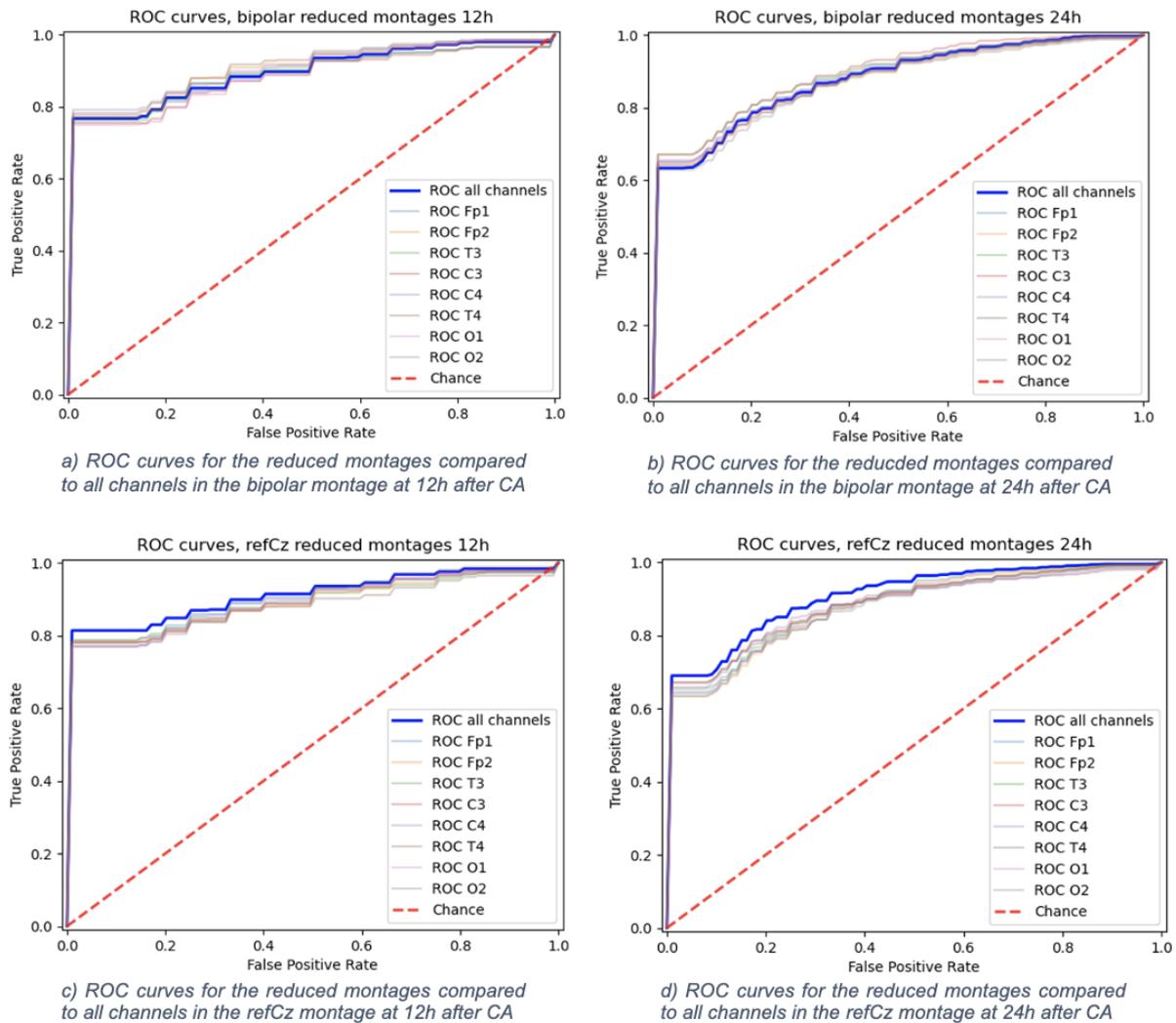


Figure A.15: Mean ROC curves over 50 iterations for all reduced montages compared to the full montages at 12h and 24h after CA. The blue line represents the full montage, and the more transparent lines all represent one reduced montage. Reduced montages are the two adjacent channels of one electrode for the bipolar montage. The performance of these reduced montages is in most cases not significantly different from the bipolar montage at both timepoints. For the refCz montage, the reduced montages are single channels in reference to Cz. These reduced montages are all outperformed by the full montage at both timepoints.

Appendix A.7. Performance LR model reduced montages

Performance metrics with 95% confidence intervals and their corresponding p-values for all reduced montages compared to the full bipolar montage at 12 hours and 24 hours after CA are shown in Table A.5 and A.6, respectively. The same metrics and values are shown in Table A.7 and A.8 for the single channel reduced montages compared to the full montage of all electrodes referenced to Cz. The comparison between the frontotemporal montage and full bipolar montage can be seen in Table A.9 and A.10 for 12 and 24 hours after CA, respectively. For all tables, coloured cells indicate significant differences compared to the baseline input of all channels. Green represents significantly higher values and red indicates significantly lower values.

Table A.5: Performance of the LR model at 12h after CA for all reduced montages compared to the full bipolar montage.

12h after resuscitation						
Input	AUC Mean (95% CI)	p-value	Se100 poor outcome Mean (95% CI)	p-value	Se95 good outcome Mean (95% CI)	p-value
All channels bipolar	0.897 (0.885-0.908)		0.767 (0.744-0.790)		0.713 (0.684-0.742)	
Bipolar, features 1:10	0.897 (0.885-0.908)	0.97	0.764 (0.741-0.788)	0.86	0.725 (0.697-0.753)	0.56
Fp1T3, Fp1C3	0.889 (0.877-0.902)	0.40	0.770 (0.747-0.792)	0.88	0.688 (0.658-0.719)	0.25
Fp2T4, Fp2C4	0.906 (0.895-0.916)	0.24	0.771 (0.748-0.794)	0.83	0.733 (0.705-0.761)	0.32
Fp1T3, T3O1	0.892 (0.880-0.904)	0.61	0.760 (0.735-0.784)	0.66	0.711 (0.680-0.741)	0.91
Fp1C3, C3O1	0.882 (0.869-0.895)	0.10	0.754 (0.730-0.777)	0.43	0.681 (0.650-0.711)	0.13
Fp2C4, C4O2	0.910 (0.899-0.920)	0.09	0.791 (0.769-0.814)	0.14	0.749 (0.722-0.776)	0.07
Fp2T4, T4O2	0.914 (0.903-0.925)	0.03	0.782 (0.758-0.806)	0.39	0.773 (0.747-0.799)	0.003
T3O1, C3O1	0.888 (0.876-0.900)	0.31	0.749 (0.725-0.773)	0.28	0.708 (0.680-0.735)	0.79
T4O2, C4O2	0.904 (0.893-0.915)	0.35	0.776 (0.754-0.799)	0.58	0.747 (0.721-0.774)	0.08

AUC: area under the curve; Se100: sensitivity at 100% specificity; Se95: sensitivity at 95% specificity; CI: confidence interval

Table A.6: Performance of the LR model at 24h after CA for all reduced montages compared to the full bipolar montage.

24h after resuscitation						
Input	AUC Mean (95% CI)	p-value	Se100 poor outcome Mean (95% CI)	p-value	Se95 good outcome Mean (95% CI)	p-value
All channels bipolar	0.879 (0.873-0.886)		0.633 (0.616-0.650)		0.496 (0.474-0.518)	
Bipolar, features 1:10	0.872 (0.864-0.879)	0.14	0.640 (0.622-0.657)	0.61	0.458 (0.433-0.482)	0.02
Fp1T3, Fp1C3	0.884 (0.877-0.891)	0.32	0.651 (0.634-0.668)	0.15	0.508 (0.484-0.532)	0.45
Fp2T4, Fp2C4	0.873 (0.866-0.880)	0.20	0.645 (0.629-0.661)	0.33	0.450 (0.427-0.474)	0.01
Fp1T3, T3O1	0.892 (0.885-0.898)	0.01	0.671 (0.655-0.688)	0.002	0.512 (0.488-0.536)	0.32
Fp1C3, C3O1	0.897 (0.891-0.904)	< 0.001	0.670 (0.653-0.687)	0.003	0.580 (0.559-0.601)	< 0.001
Fp2C4, C4O2	0.878 (0.871-0.885)	0.83	0.649 (0.633-0.666)	0.18	0.484 (0.460-0.508)	0.47
Fp2T4, T4O2	0.875 (0.867-0.882)	0.36	0.641 (0.624-0.657)	0.54	0.481 (0.458-0.504)	0.36
T3O1, C3O1	0.884 (0.877-0.891)	0.31	0.655 (0.639-0.672)	0.07	0.525 (0.502-0.547)	0.07
T4O2, C4O2	0.867 (0.860-0.874)	0.02	0.627 (0.610-0.645)	0.64	0.449 (0.423-0.474)	0.01

AUC: area under the curve; Se100: sensitivity at 100% specificity; Se95: sensitivity at 95% specificity; CI: confidence interval

Table A.7: Performance of the LR model at 12h after CA for all reduced montages compared to the full refCz montage.

12h after resuscitation						
Input	AUC Mean (95% CI)	p-value	Se100 poor outcome Mean (95% CI)	p-value	Se95 good outcome Mean (95% CI)	p-value
All channels refCz	0.912 (0.902-0.923)		0.814 (0.793-0.834)		0.745 (0.717-0.773)	
RefCz, features 1:10	0.900 (0.889-0.911)	0.11	0.782 (0.761-0.803)	0.04	0.719 (0.691-0.748)	0.20
Fp1Cz	0.898 (0.886-0.910)	0.07	0.781 (0.759-0.803)	0.04	0.718 (0.689-0.748)	0.19
Fp2Cz	0.889 (0.876-0.902)	0.01	0.781 (0.759-0.804)	0.04	0.697 (0.666-0.727)	0.02
T3Cz	0.893 (0.880-0.905)	0.02	0.788 (0.767-0.810)	0.09	0.699 (0.669-0.729)	0.03
C3Cz	0.892 (0.881-0.904)	0.01	0.779 (0.758-0.801)	0.02	0.691 (0.661-0.720)	0.01
C4Cz	0.895 (0.884-0.906)	0.03	0.769 (0.746-0.791)	0.004	0.703 (0.673-0.732)	0.04
T4Cz	0.883 (0.870-0.896)	< 0.001	0.785 (0.763-0.807)	0.06	0.671 (0.639-0.704)	< 0.001
O1Cz	0.892 (0.881-0.903)	0.01	0.770 (0.748-0.792)	0.01	0.699 (0.670-0.727)	0.02
O2Cz	0.896 (0.885-0.908)	0.04	0.771 (0.749-0.793)	0.01	0.724 (0.698-0.750)	0.27

AUC: area under the curve; Se100: sensitivity at 100% specificity; Se95: sensitivity at 95% specificity; CI: confidence interval

Table A.8: Performance of the LR model at 24h after CA for all reduced montages compared to the full refCz montage.

24h after resuscitation						
Input	AUC Mean (95% CI)	p-value	Se100 poor outcome Mean (95% CI)	p-value	Se95 good outcome Mean (95% CI)	p-value
All channels refCz	0.910 (0.904-0.916)		0.690 (0.673-0.706)		0.597 (0.573-0.622)	
RefCz, features 1:10	0.902 (0.896-0.908)	0.08	0.685 (0.668-0.702)	0.70	0.565 (0.540-0.591)	0.07
Fp1Cz	0.884 (0.877-0.891)	< 0.001	0.642 (0.625-0.659)	< 0.001	0.541 (0.517-0.565)	0.001
Fp2Cz	0.880 (0.873-0.886)	< 0.001	0.634 (0.618-0.651)	< 0.001	0.510 (0.485-0.534)	< 0.001
T3Cz	0.883 (0.876-0.890)	< 0.001	0.654 (0.637-0.671)	0.003	0.491 (0.466-0.517)	< 0.001
C3Cz	0.887 (0.880-0.894)	< 0.001	0.671 (0.655-0.687)	0.11	0.496 (0.471-0.521)	< 0.001
C4Cz	0.884 (0.876-0.891)	< 0.001	0.659 (0.640-0.678)	0.01	0.459 (0.431-0.486)	< 0.001
T4Cz	0.887 (0.879-0.894)	< 0.001	0.671 (0.654-0.688)	0.13	0.505 (0.478-0.532)	< 0.001
O1Cz	0.876 (0.869-0.883)	< 0.001	0.646 (0.629-0.664)	< 0.001	0.457 (0.428-0.486)	< 0.001
O2Cz	0.872 (0.865-0.879)	< 0.001	0.435 (0.408-0.462)	< 0.001	0.435 (0.408-0.462)	< 0.001

AUC: area under the curve; Se100: sensitivity at 100% specificity; Se95: sensitivity at 95% specificity; CI: confidence interval

Table A.9: Performance of the LR model at 12h after CA for the frontotemporal montage compared to the full bipolar montage.

12h after resuscitation						
Input	AUC Mean (95% CI)	p-value	Se100 poor outcome Mean (95% CI)	p-value	Se95 good outcome Mean (95% CI)	p-value
All channels bipolar	0.897 (0.885-0.908)		0.767 (0.744-0.790)		0.713 (0.684-0.742)	
FT montage	0.902 (0.890-0.913)	0.54	0.801 (0.779-0.823)	0.04	0.726 (0.696-0.756)	0.53

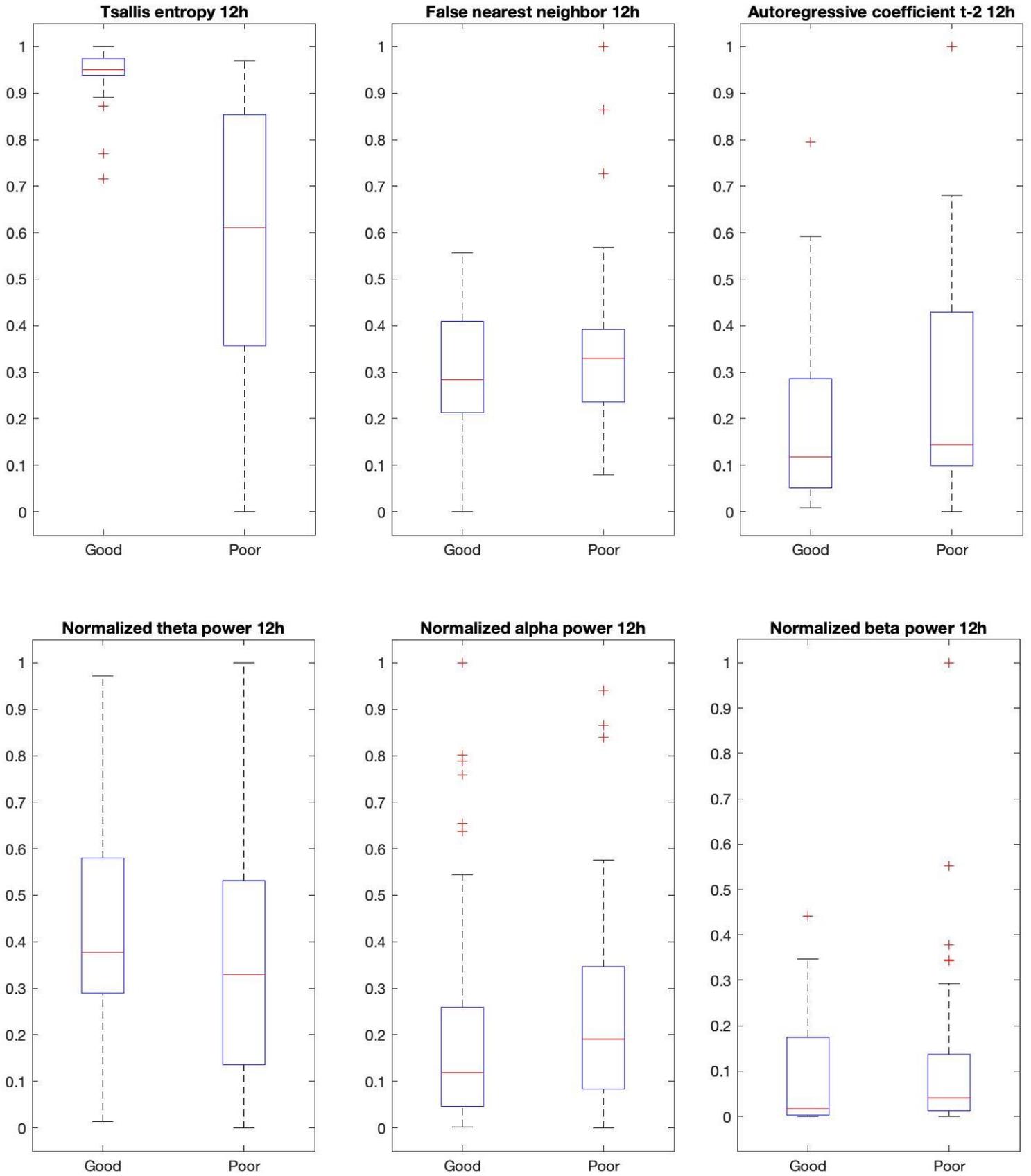
AUC: area under the curve; Se100: sensitivity at 100% specificity; Se95: sensitivity at 95% specificity; CI: confidence interval

Table A.10: Performance of the LR model at 24h after CA for the frontotemporal montage compared to the full bipolar montage.

24h after resuscitation						
Input	AUC Mean (95% CI)	p-value	Se100 poor outcome Mean (95% CI)	p-value	Se95 good outcome Mean (95% CI)	p-value
All channels bipolar	0.879 (0.873-0.886)		0.633 (0.616-0.650)		0.496 (0.474-0.518)	
FT montage	0.896 (0.889-0.902)	< 0.001	0.682 (0.665-0.698)	< 0.001	0.521 (0.496-0.545)	0.14

AUC: area under the curve; Se100: sensitivity at 100% specificity; Se95: sensitivity at 95% specificity; CI: confidence interval

Appendix A.8. Boxplots feature values



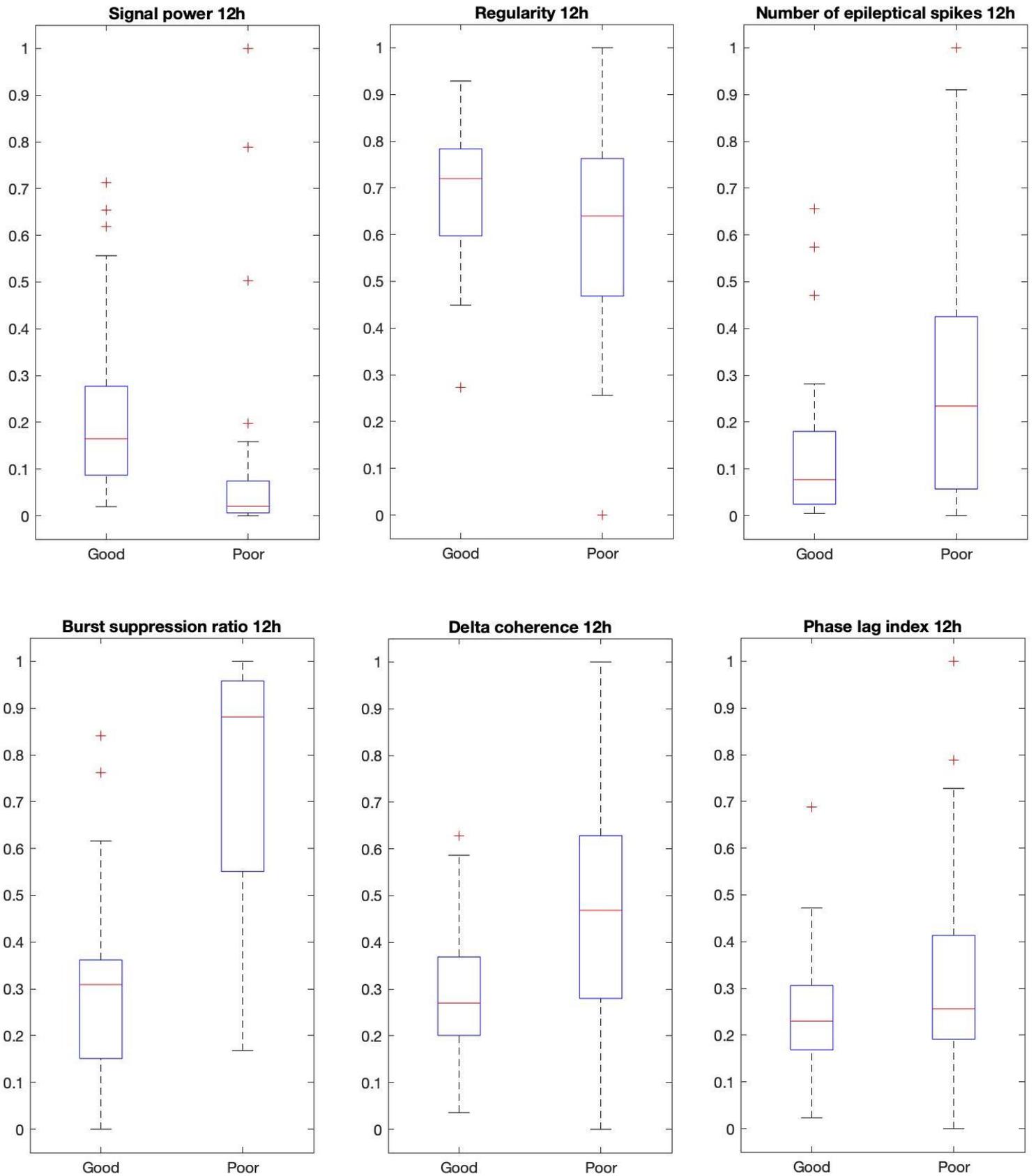
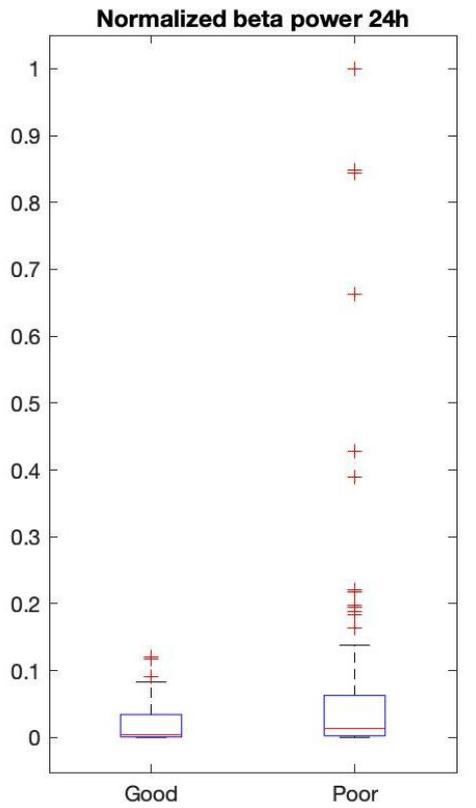
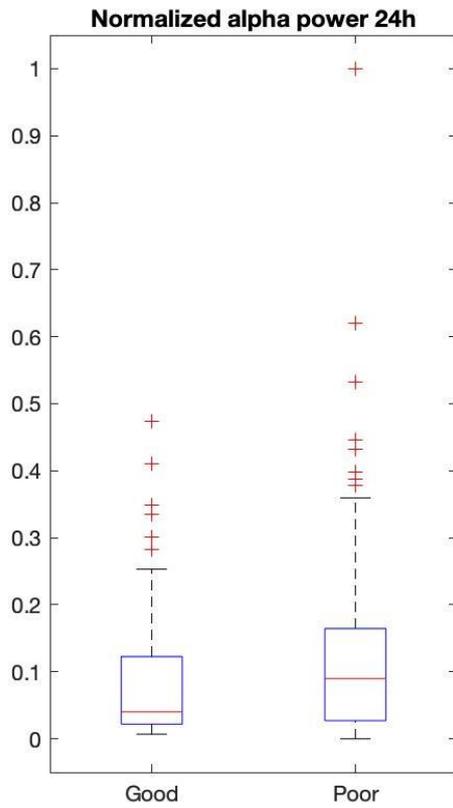
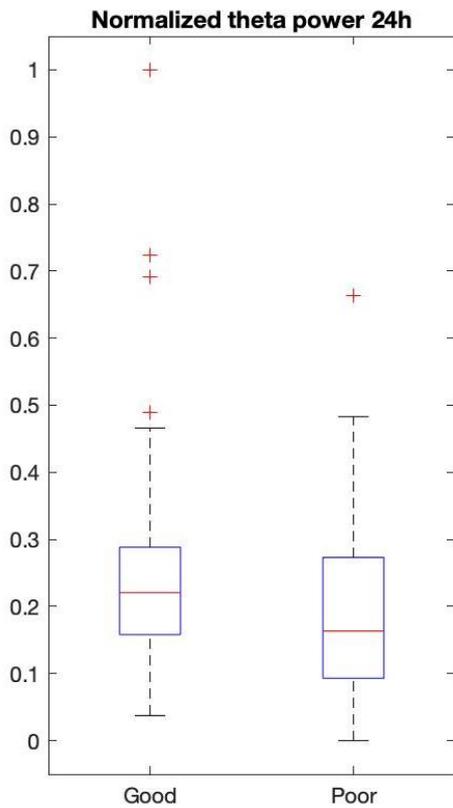
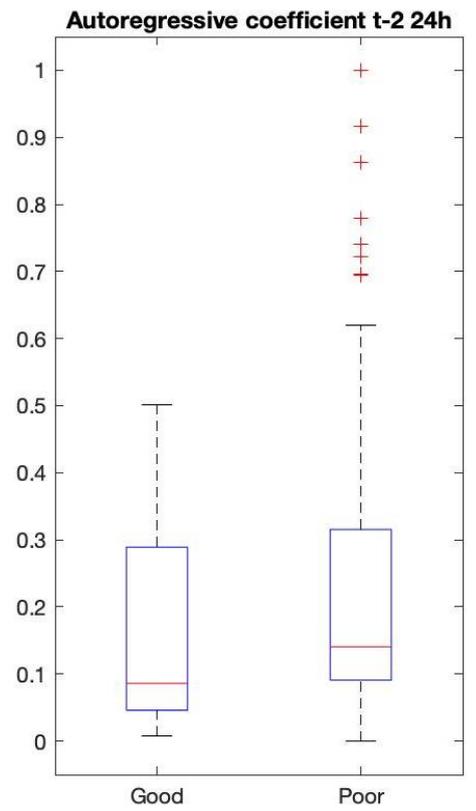
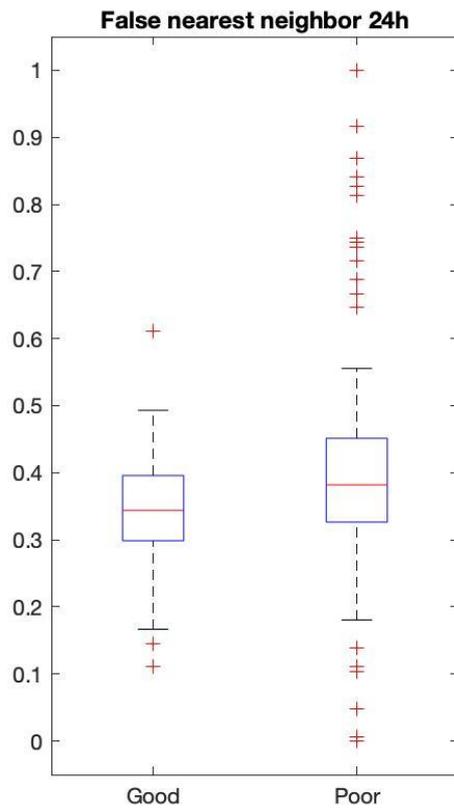
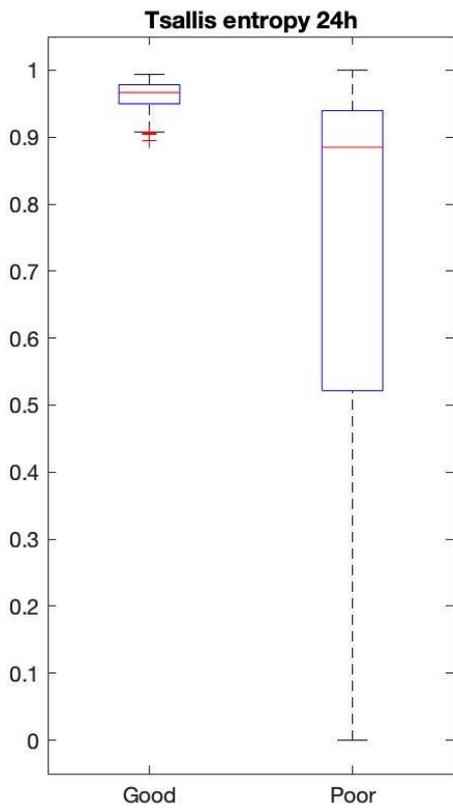


Figure A.16: Boxplots of the feature values at 12 hours after CA for all patients with good and poor neurological outcome. The red central mark indicates the median of the data, the bottom and top edge of the box represent 25 and 75 percentiles respectively. The whiskers extend to the most extreme data points, outliers excluded. Outliers are represented by the red plus signs. The features Tsallis entropy, normalized theta power, signal power and regularity show a lower median value in patients with poor outcome compared to patients with good outcome. All other features show higher median values for patients with poor outcome.



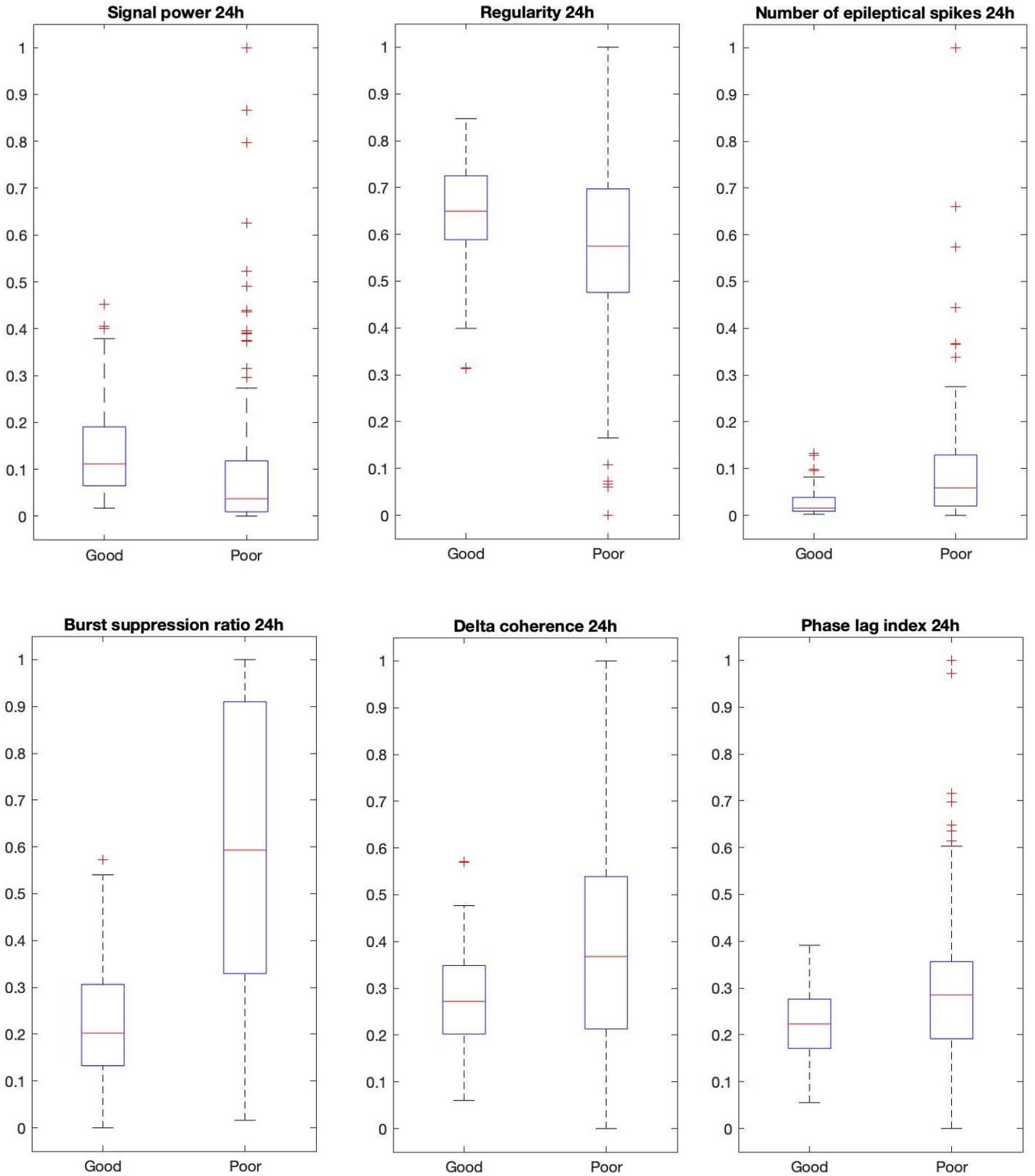


Figure A.17: Boxplots of the feature values at 24 hours after CA for all patients with good and poor neurological outcome. The red central mark indicates the median of the data, the bottom and top edge of the box represent 25 and 75 percentiles respectively. The whiskers extend to the most extreme data points, outliers excluded. Outliers are represented by the red plus signs. The features Tsallis entropy, normalized theta power, signal power and regularity show a lower median value in patients with poor outcome compared to patients with good outcome. All other features show higher median values for patients with poor outcome.

Appendix A.9. Performance and trained weights L1 regularization

The mean of the absolute values of the trained weights and their corresponding performance metrics for different values of λ when performing L1 regularization are shown in Table A.11 and A.12 for 12 hours after CA and in Table A.13 and A.14 for 24 hours after CA.

Table A.11: Trained weights for all input channels when applying L1 regularization with different values of λ at 12h after CA.

Trained weights Mean of absolute values	12h after resuscitation							
	Channels							
Input	Fp1	Fp2	T3	C3	C4	T4	O1	O2
No regularizer	0,45	0,49	1,19	0,67	0,33	0,91	0,56	0,56
L1, $\lambda = 0.0005$	0,44	0,52	1,20	0,76	0,30	0,99	0,66	0,55
L1, $\lambda = 0.001$	0,41	0,48	1,10	0,74	0,25	0,89	0,55	0,52
L1, $\lambda = 0.005$	0,29	0,19	0,81	0,55	0,12	0,34	0,25	0,16
L1, $\lambda = 0.01$	0,14	0,09	0,57	0,39	0,08	0,04	0,03	0,05
L1, $\lambda = 0.05$	0,00	0,00	0,30	0,10	0,00	0,00	0,00	0,00
L1, $\lambda = 0.1$	0,00	0,00	0,03	0,01	0,00	0,00	0,00	0,00
L1, $\lambda = 0.5$	0,00	0,00	0,01	0,01	0,00	0,00	0,00	0,00

Table A.12: Corresponding model performance of the LR model using L1 regularization for different values of λ at 12h after CA.

12h after resuscitation			
Input	AUC	Se100 poor	Se95 good
No regularizer	0,871	0,726	0,646
L1, $\lambda = 0.0005$	0,858	0,696	0,641
L1, $\lambda = 0.001$	0,870	0,732	0,650
L1, $\lambda = 0.005$	0,878	0,729	0,664
L1, $\lambda = 0.01$	0,873	0,743	0,674
L1, $\lambda = 0.05$	0,886	0,779	0,687
L1, $\lambda = 0.1$	0,893	0,781	0,691
L1, $\lambda = 0.5$	0,795	0,666	0,607

Table A.13: Trained weights for all input channels when applying L1 regularization with different values of λ at 24h after CA.

Trained weights Mean of absolute values	24h after resuscitation							
	Channels							
Input	Fp1	Fp2	T3	C3	C4	T4	O1	O2
No regularizer	1,20	0,71	0,25	1,44	0,30	0,59	0,34	0,99
L1, $\lambda = 0.0005$	1,10	0,56	0,20	1,41	0,25	0,47	0,29	0,94
L1, $\lambda = 0.001$	1,01	0,43	0,14	1,27	0,19	0,41	0,25	0,88
L1, $\lambda = 0.005$	0,46	0,01	0,03	1,04	0,01	0,17	0,03	0,25
L1, $\lambda = 0.01$	0,28	0,00	0,01	0,88	0,00	0,09	0,01	0,01
L1, $\lambda = 0.05$	0,00	0,00	0,00	0,15	0,00	0,00	0,00	0,00
L1, $\lambda = 0.1$	0,00	0,00	0,00	0,10	0,00	0,00	0,01	0,00
L1, $\lambda = 0.5$	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01

Table A.14: Corresponding model performance of the LR model using L1 regularization for different values of λ at 24h after CA.

24h after resuscitation			
Input	AUC	Se100 poor	Se95 good
No regularizer	0,837	0,643	0,304
L1, $\lambda = 0.0005$	0,835	0,640	0,305
L1, $\lambda = 0.001$	0,841	0,643	0,317
L1, $\lambda = 0.005$	0,846	0,644	0,332
L1, $\lambda = 0.01$	0,849	0,654	0,322
L1, $\lambda = 0.05$	0,853	0,654	0,337
L1, $\lambda = 0.1$	0,853	0,650	0,351
L1, $\lambda = 0.5$	0,847	0,649	0,339

Appendix A.10. LR performance compared with literature

Table A.15: Performance of outcome prediction models from earlier studies in the AMC and multiple other studies at 12 hours after CA.

12h after CA			
Model (montage)	AUC of ROC Mean (95% CI)	Sens (at Spec) for poor outcome prediction Mean (95% CI)	Sens (at Spec) for good outcome prediction Mean (95% CI)
Current model			
Current LR (9 electrodes, bipolar)	0.90 (0.89-0.91)	0.77 (0.74-0.79) (100%)	0.71 (0.68-0.74) (95%)
Current LR (9 electrodes, refCz)	0.91 (0.90-0.92)	0.81 (0.79-0.83) (100%)	0.75 (0.72-0.77) (95%)
Current LR (4 electrodes, frontotemporal)	0.90 (0.89-0.91)	0.80 (0.78-0.82) (100%)	0.73 (0.70-0.76) (95%)
Earlier studies AMC			
CRI (9 electrodes, bipolar) ⁶²	0.80	0.09 (0.01-0.29) (100%)	0.73 (0.39-0.94) (73%)
CRI (3 electrodes, central) ⁶²	0.83	0.22 (0.09-0.42) (100%)	0.90 (0.73-0.98) (70%)
rCRI (9 electrodes, bipolar) ³⁸	0.87	0.88 (71%)	0.00 (100%)
rCRI (3 electrodes, central) ³⁸	0.91	0.65 (100%)	0.21 (100%)
LR (9 electrodes, bipolar) ²⁴	0.90 (0.89-0.91)	0.78 (0.76-0.80) (100%)	0.30 (0.27-0.32) (95%)
LSTM (9 electrodes, bipolar) ²⁴	0.90 (0.89-0.91)	0.79 (0.76-0.81) (100%)	0.30 (0.28-0.33) (95%)
Literature			
CRI (21 electrodes, bipolar) ⁸	0.92	0.56 (0.41-0.70) (100%)	0.63 (0.46-0.77) (94%)
rCRI (21 electrodes) ²⁰	0.94	0.66 (0.65-0.78) (100%)	0.72 (0.61-0.85) (95%)
CNN (21 electrodes, bipolar) ²³	0.87 (0.87-0.88)	0.42 (0.36-0.48) (100%)	0.48 (0.45-0.51) (95%)
CNN (19 electrodes, single channels) ²¹	0.89 (0.78-0.96)	0.78 (89%)	

CA: cardiac arrest; AUC: area under the curve; ROC: receiver operating characteristic; Se100: sensitivity at 100% specificity; Se95: sensitivity at 95% specificity; CI: confidence interval; LR: logistic regression; (r)CRI: (revised) cerebral recovery index; LSTM: long short-term memory recurrent neural network; CNN: convolutional neural network.

Table A.16: Performance of outcome prediction models from earlier studies in the AMC and multiple other studies at 24 hours after CA.

24h after CA			
Model (montage)	AUC of ROC Mean (95% CI)	Sens (at Spec) for poor outcome prediction Mean (95% CI)	Sens (at Spec) for good outcome prediction Mean (95% CI)
Current model			
Current LR (9 electrodes, bipolar)	0.88 (0.87-0.89)	0.63 (0.62-0.65) (100%)	0.50 (0.47-0.52) (95%)
Current LR (9 electrodes, refCz)	0.91 (0.90-0.92)	0.69 (0.67-0.71) (100%)	0.60 (0.57-0.62) (95%)
Current LR (4 electrodes, frontotemporal)	0.90 (0.89-0.90)	0.68 (0.67-0.70) (100%)	0.52 (0.50-0.55) (95%)
Earlier studies AMC			
CRI (9 electrodes, bipolar) ⁶²	0.75	0.54 (0.36-0.70) (100%)	0.55 (0.23-0.83) (78%)
CRI (3 electrodes, central) ⁶²	0.70	0.39 (0.28-0.52) (98%)	0.82 (0.69-0.92) (58%)
rCRI (9 electrodes, bipolar) ³⁸	0.74	0.79 (50%)	0.17 (95%)
rCRI (3 electrodes, central) ³⁸	0.86	0.34 (100%)	0.45 (97%)
LR (9 electrodes, bipolar) ²⁴	0.88 (0.88-0.89)	0.67 (0.66-0.69) (100%)	0.56 (0.53-0.58) (95%)
LSTM (9 electrodes, bipolar) ²⁴	0.90 (0.90-0.91)	0.68 (0.66-0.70) (100%)	0.44 (0.42-0.46) (95%)
Literature			
CRI (21 electrodes, bipolar) ⁸	0.90	0.65 (0.51-0.77) (94%)	0.58 (0.43-0.71) (93%)
rCRI (21 electrodes) ²⁰	0.88	0.60 (0.51-0.75) (100%)	0.40 (0.30-0.51) (95%)
CNN (21 electrodes, bipolar) ²³	0.90 (0.90-0.91)	0.57 (0.54-0.60) (100%)	0.33 (0.30-0.36) (95%)

CA: cardiac arrest; AUC: area under the curve; ROC: receiver operating characteristic; Se100: sensitivity at 100% specificity; Se95: sensitivity at 95% specificity; CI: confidence interval; LR: logistic regression; (r)CRI: (revised) cerebral recovery index; LSTM: long short-term memory recurrent neural network; CNN: convolutional neural network.