Digitale Anfallsdetektion und Qualitätsbewertung in der Elektrokonvulsionstherapie

Dissertation

zur Erlangung des Doktorgrades (Dr. med.)

der Medizinischen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität

Bonn

Max Kayser

aus Luxemburg / Großherzogtum Luxemburg

2026

Angefertigt mit der Genehmigung

der Medizinischen Fakultät der Universität Bonn

1. Gutachterin: Univ.-Prof. Dr. med. Alexandra Philipsen
2. Gutachter: Prof. Dr. rer. medic. Stephan Jonas

Tag der mündlichen Prüfung: 7. Juli 2026

Aus der Klinik für Psychiatrie und Psychotherapie

Inhaltsverzeichnis

Abkürzungsverzeichnis		**7**
1.	**Deutsche Zusammenfassung**	**9**
1.1	Einleitung	9
1.1.1	Hintergrund und klinische Bedeutung der Elektrokonvulsionstherapie	9
1.1.2	Herausforderungen bei der EEG-basierten Anfallsbewertung	10
1.1.3	Quantitative EEG-Parameter zur Krampfqualitätsbewertung	11
1.1.4	Potenziale digitaler Analysemethoden und Künstlicher Intelligenz	12
1.1.5	Forschungsfragen	13
1.1.6	Zielsetzung	13
1.2	Material und Methoden	14
1.2.1	Ethik und regulatorische Rahmenbedingungen	15
1.2.2	Elektrokonvulsionstherapie	15
1.2.2.1	Patientenkollektiv und Indikation	15
1.2.2.2	Technische Durchführung und Stimulationsparameter	16
1.2.2.3	Anästhesie und Monitoring	17
1.2.2.4	Klinische Verlaufsdokumentation	18
1.2.3	Datenerfassung und EEG-Aufzeichnung	18
1.2.4	Datenaufbereitung	19
1.2.5	Anfallserkennung	20
1.2.5.1	Maschinelles Lernen	21
1.2.5.2	Merkmalsextraktion	21

1.2.5.3	Balancierung und Aufteilung der Daten	23
1.2.5.4	Aufbau des digitalen Analyse-Frameworks	24
1.2.5.5	Datenvorverarbeitung	24
1.2.5.6	Algorithmen des maschinellen Lernens	24
1.2.5.7	Training der Klassifikationsmodelle	26
1.2.5.8	Evaluation und Leistungsbewertung der Modelle	26
1.2.5.9	Modellvergleichsanalyse	29
1.2.5.10	Anfallserkennung	30
1.2.6	Berechnung der Qualitätsparameter	32
1.2.6.1	Krampfdauer	33
1.2.6.2	Postiktaler Suppressionindex	34
1.2.6.3	Konvulsions-Energie-Index	35
1.2.6.4	Maximal anhaltende Leistung	35
1.2.6.5	Zeit bis zur maximalen Leistung	36
1.2.6.6	Frühe iktale Amplitude	36
1.2.6.7	Mittlere iktale Amplitude	37
1.2.6.8	Maximale interhemisphärische Kohärenz	38
1.2.6.9	Statistische Analyse	39
1.3	Ergebnisse	40
1.3.1	Datensatz und klinische Charakteristika	40
1.3.2	Segmentbasierte Anfallserkennung	42
1.3.2.1	Merkmalsanalyse	43
1.3.2.2	Hyperparameteroptimierung	45

1.3.2.3	Modellvergleichsanalyse	47
1.3.3	Analyse der Qualitätsparameter	50
1.4	Diskussion	59
1.4.1	Zentrale Befunde	59
1.4.2	Interpretation der Ergebnisse	59
1.4.2.1	Datensatz und klinische Charakteristika	60
1.4.2.2	Implementierung und Reproduzierbarkeit	61
1.4.2.3	Stichprobentiefe und Auswertungsschnitt	62
1.4.2.4	Merkmalsraum	62
1.4.2.5	Hyperparameteroptimierung	63
1.4.2.6	Leistungsbewertung der Modelle	64
1.4.2.7	Qualitätsparameter	65
1.4.3	Limitationen	67
1.4.3.1	Datengrundlage und Kohorten-Struktur	67
1.4.3.2	Setting und Gerätetechnik	67
1.4.3.3	Analyseverfahren und Auswertung	68
1.4.3.4	Referenz- und Definitionsabhängigkeit	70
1.4.4	Beantwortung der Fragestellungen	71
1.4.5	Klinische Implikationen	72
1.4.6	Ausblick	74
1.4.7	Schlussfolgerung	76
1.5	Zusammenfassung	76
1.6	**Literaturverzeichnis der deutschen Zusammenfassung**	77

2. Veröffentlichung 90

3. Erklärung zum Eigenanteil 91

4. Danksagung 93

Abkürzungsverzeichnis

ASEI	Konvulsions-Energie-Index (engl. Average Seizure Energy Index)
AUC	Fläche unter der Kurve (engl. *Area Under the Curve*)
CGI	Klinischer Gesamteindruck (engl. *Clinical Global Impression*)
CI	Konfidenzintervall (engl. *Confidence Interval*)
COH	Maximale interhemisphärische Kohärenz (engl. *Maximum Sustained Coherence*)
DT	Entscheidungsbaum (engl. *Decision Tree*)
EEG	Elektroenzephalogramm
EIA	Frühe iktale Amplitude (engl. *Early-Ictal Amplitude*)
EKT	Elektrokonvulsionstherapie
FN	Falsch negativ (engl. *False Negative*)
FP	Falsch positiv (engl. *False Positive*)
FPR	Falsch-Positiv-Rate (engl. *False Positive Rate*)
GBC	Gradient-Boosting-Klassifikator (engl. *Gradient Boosting Classifier*)
ICC	Intraklassen-Korrelationskoeffizient (engl. *Intraclass Correlation Coefficient*)
ICD	Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme (engl. *International Statistical Classification of Diseases and Related Health Problems*)
KNN	k-Nächste-Nachbarn Klassifikator (engl. *K-Nearest Neighbours Classifier*)

MCC	Matthews Korrelationskoeffizient (engl. *Matthews Correlation Coefficient*)
MIA	Mittlere iktale Amplitude (engl. *Mid-ictal Amplitude*)
ML	Maschinelles Lernen (engl. *Machine Learning*)
MSP	Maximal anhaltende Leistung (engl. *Maximum Sustained Power*)
MSE	Standardfehler des Mittelwerts
MWU	Mann-Whitney-U-Test
PPV	Positiver prädiktiver Wert (engl. *Positive Predictive Value*)
PSD	Leistungsspektraldichte (engl. *Power Spectral Density*)
PSI	Postiktaler Suppressionindex (engl. *Postictal Suppression Index*)
RF	Random Forest
ROC	Grenzwertoptimierung (engl. *Receiver Operating Characteristic*)
SD	Standardabweichung
SE	Krampfdauer
SVC	Stützvektor-Klassifikator (engl. *Support Vector Classifier*)
TN	Richtig negativ (engl. *True Negative*)
TP	Richtig positiv (engl. *True Positive*)
TPR	Richtig-Positiv-Rate (engl. *True Positive Rate*)
TTPP	Zeit bis zur maximalen Leistung (engl. *Time To Peak Power*)
UKB	Universitätsklinikum Bonn
UKE	Universitätsklinikum Hamburg-Eppendorf

1. Deutsche Zusammenfassung

Hinweis zur Sprache

Diese Arbeit verwendet primär geschlechtsneutrale Formulierungen. In Einzelfällen, in denen dies nicht möglich oder stark schwerfällig wäre, kommt aus Gründen der Lesbarkeit das generische Maskulinum zur Anwendung. Alle Personenbezeichnungen gelten sofern nicht anders vermerkt, für Menschen aller Geschlechter.

1.1 Einleitung

1.1.1 Hintergrund und klinische Bedeutung der Elektrokonvulsionstherapie

Die Elektrokonvulsionstherapie (EKT) ist eines der am längsten etablierten biologischen Behandlungsverfahren der Psychiatrie und gehört weiterhin zu den wirksamsten Akuttherapien bei schweren affektiven Störungen (Weiner und Reti, 2017). Ihre Ursprünge reichen bis in die 1930er Jahre zurück, als Ladislaus von Meduna erstmals durch pharmakologisch induzierte Krampfanfälle therapeutische Effekte bei psychiatrischen Erkrankungen beobachtete (Fink, 1984). Kurz darauf etablierten Cerletti und Bini die elektrische Auslösung generalisierter Krampfanfälle beim Menschen und legten damit die Grundlage der modernen EKT (Accornero, 1988). In ihrer heutigen Form wird die EKT unter Vollnarkose und Muskelrelaxation durchgeführt und gilt als sichere und hoch effektive Behandlungsoption bei therapieresistenter Depression, bipolarer Störung und Schizophrenie (Espinoza und Kellner, 2022; Kellner et al., 2009).

Obgleich der klinische Nutzen der EKT seit Jahrzehnten belegt ist (Pagnin et al., 2004), bleiben zentrale Aspekte ihrer neurobiologischen Wirkweise bis heute unvollständig verstanden (Espinoza und Kellner, 2022; Farzan et al., 2014). So zeigen Studien komplexe Veränderungen in Neurotransmittersystemen, neurotrophen Faktoren und neuronalen Netzwerken (Perera et al., 2007; Sanacora et al., 2003), die mit den antidepressiven Effekten korrespondieren könnten. Kern der Behandlung ist die Induktion eines generalisierten tonisch-klonischen Krampfanfalls, dessen Qualität in engem Zusammenhang mit dem Therapieerfolg steht (Fink, 2001; Hoppe und Scriba, 2003; Janouschek et al., 2020).

1.1.2 Herausforderungen bei der EEG-basierten Anfallsbewertung

Das Elektroenzephalogramm (EEG) ist ein zentrales Instrument zur Überwachung und Beurteilung der EKT (Christensen und Koldbæk, 1982; Weiner et al., 1991). Früh wurde deutlich, dass EEG-Analysen neben der Darstellung von Krampfaktivität auch wichtige Informationen zur Therapiequalität bereitstellen (Fink, 1957; Mayur, 2006). Zahlreiche Studien untersuchten die Relevanz des EEGs für die Bewertung der Therapieantwort, insbesondere im Zusammenhang mit Veränderungen der EEG-Signalverläufe vor, während und nach der Behandlung (Francis-Taylor et al., 2020; Olbrich und Arns, 2013; Sackeim, 1996; Scangos et al., 2019; Simon et al., 2021). Die Krampfdauer gilt seit langem als potenzieller Erfolgsindikator (Gillving et al., 2024; Sackeim et al., 1991), und sollte mindestens 25–30 Sekunden betragen (Beyer et al., 1998; Grözinger et al., 2013). Eine zu kurze Krampfdauer kann mit einem unzureichenden Therapieerfolg assoziiert sein, während längere Anfälle nicht zwangsläufig zu besseren klinischen Ergebnissen führen (Swartz et al., 1994; Warmflash et al., 1987).

Die manuelle Bestimmung der Anfallsdauer ist jedoch fehleranfällig und weist eine hohe Inter- und Intra-Rater-Variabilität auf (Guze et al., 1989; Kramer et al., 1989; Krystal et al., 1993; Ries, 1985; Warmflash et al., 1987). Um diese Limitationen zu überwinden, wurden gerätebasierte Methoden zur automatisierten Anfallserkennung entwickelt (Krystal und Weiner, 1995; Swartz et al., 1994). Allerdings gibt es nicht ausreichend Studien, die die Validität dieser automatisierten Verfahren ausreichend belegen (Scott, 2007). Zudem kann die Zuverlässigkeit der Anfallserkennung durch Signalartefakte, eine schwache postiktale Suppression oder ein schrittweises Abklingen der Anfallsaktivität erheblich beeinträchtigt werden (Krystal und Weiner, 1995). In einer Untersuchung konnte eine computergestützte Methode bei fast einem Drittel der EKT-Behandlungen das Anfallsende nicht zuverlässig bestimmen (Rosenquist et al., 1998). Abseits der geräteinternen Algorithmen liegen bislang nur vereinzelt Studien zu alternativen Ansätzen zur automatisierten Anfallserkennung vor. Die wenigen vorhandenen wissenschaftlichen Arbeiten befassen sich unter anderem mit der Berechnung der fraktalen Dimension (Gangadhar et al., 1995), der Sample Entropy (Azuma et al., 2020), sowie der Anwendung von Wavelet Transformationen zur Bestimmung von Anfallsdauer und Anfallsendpunkt (Yoo et al., 2012). Diese Ver-

fahren eröffnen zusätzliche Möglichkeiten für eine objektive Bewertung der Krampfqualität. Ihre klinische Anwendbarkeit ist derzeit jedoch begrenzt und erfordert weiterführende Validierungsstudien. In der klinischen Praxis existieren bislang keine etablierten computergestützten Alternativen zur Detektion von durch EKT induzierten Krampfanfällen.

1.1.3 Quantitative EEG-Parameter zur Krampfqualitätsbewertung

Die Krampfdauer allein erlaubt keine differenzierte und umfassende Bewertung der Krampfqualität. Daher müssen weitere potenzielle Marker berücksichtigt werden (Krystal et al., 1993; Nobler et al., 1993). Damit ist die ausgeprägte Reduktion der kortikalen Aktivität im Anschluss an den Krampf gemeint. Mehrere Studien zeigen, dass eine stärkere postiktale Unterdrückung mit einem besseren therapeutischen Ansprechen assoziiert ist (Azuma et al., 2007; Suppes et al., 1996; Weiner et al., 1991). Auch Amplituden- und Frequenzverläufe werden untersucht, um Rückschlüsse auf die Krampfqualität (Francis-Taylor et al., 2020; Krystal et al., 1995) oder das Ansprechen auf die Therapie zu ziehen (Simon et al., 2021). Untersuchungen zeigen zudem, dass die Intensität der interhemisphärischen Synchronizität eine Rolle für den Therapieerfolg spielen könnte (Roemer et al., 1990; Ten Doesschate et al., 2019).

Technische Parameter können die EEG-Signale und damit die Interpretation der Krampfqualität verändern. Dazu gehören Stimulationsprotokolle (Niemantsverdriet et al., 2011; Nobler et al., 1993; Squire und Zouzounis, 1986; Swartz, 2000), die Elektrodenplatzierung (Krystal et al., 1993; Nobler et al., 1993) und die gewählten Anästhesieverfahren (Yalcin et al., 2012). Diese Faktoren beeinflussen sowohl die Krampfschwelle als auch die Wirksamkeit der Behandlung (Delva et al., 2000; Sackeim et al., 1991). Studien zeigen, dass die Krampfschwelle interindividuell variiert und durch unterschiedliche Einflussfaktoren moduliert wird. Dazu zählen unter anderem Alter, begleitende Medikation und das eingesetzte Anästhetikum (Aytuluk et al., 2019; van Waarde et al., 2013).

Um das Potenzial quantitativer EEG-Parameter ausschöpfen zu können, ist eine standardisierte und konsistente Erfassung erforderlich. In der klinischen Praxis wurde die Durchführung und Dokumentation der EKT über viele Jahre hinweg uneinheitlich gehandhabt. Dies erschwerte die Vergleichbarkeit wissenschaftlicher Studien erheblich (Haesebaert et

al., 2020; Prudic et al., 2001). Erst mit der Einführung digitaler Dokumentationssysteme wie GENET GPD konnte eine standardisierte und systematische Erfassung relevanter EEG-Parameter etabliert werden (Freundlieb et al., 2023; Kiebs und Freundlieb, 2021). Diese Entwicklung ermöglicht es, große EKT Datensätze für weiterführende Analysen nutzbar zu machen und damit die Bewertung der Therapieeffekte objektiver zu gestalten (Lundin et al., 2024; Zandi et al., 2022).

1.1.4 Potenziale digitaler Analysemethoden und Künstlicher Intelligenz

Mit dem Fortschritt der Digitalisierung gewinnen datengetriebene Methoden zunehmend an Bedeutung in der psychiatrischen Forschung. Insbesondere Methoden des maschinellen Lernens (engl. *machine learning*; ML) als Teilbereich der künstlichen Intelligenz werden verstärkt eingesetzt. Sie ermöglichen es, komplexe Zusammenhänge in hochdimensionalen klinischen und neurobiologischen Daten abzubilden und spielen eine wichtige Rolle in der personalisierten Medizin. In verschiedenen Anwendungsbereichen konnten prädiktive Modelle entwickelt werden, die individuelle Therapieantworten und Verläufe mit hoher Genauigkeit vorhersagen (Bzdok und Meyer-Lindenberg, 2018; Itani und Rossignol, 2020; Lundin und Menkes, 2021; Triantafyllidis und Tsanas, 2019).

Auch im Kontext der EKT wurden ML-Modelle zur Prädiktion des Therapieansprechens entwickelt (Lundin et al., 2024). Diese Modelle basieren häufig auf klinischen Routinedaten und erlauben die Abschätzung von Behandlungsergebnissen sowie die Entwicklung personalisierter Therapieansätze (Abbott et al., 2016; de Vreede et al., 2005). Jenseits klinischer Routinedaten konnte gezeigt werden, dass strukturelle und funktionelle MRT-Daten eine zusätzliche Grundlage für die prädiktive Modellierung des Therapieansprechens bieten (Gärtner et al., 2021; Redlich et al., 2016; Sun et al., 2020). Neben der Vorhersage klinischer Outcomes liegen erste Arbeiten zur automatisierten Bewertung der Krampfqualität vor. Ein Beispiel ist ein Ansatz, der Entscheidungsbäume und Fuzzy Inferenz Systeme kombiniert, um Merkmale der Krampfqualität algorithmisch zu klassifizieren (Al-Kaysi et al., 2018).

1.1.5 Forschungsfragen

Trotz vielversprechender Fortschritte in der automatisierten EEG-Analyse bleiben die präzise Erfassung und Bewertung der Krampfqualität eine zentrale Herausforderung. Insbesondere fehlt es an einem standardisierten, klinisch validierten Verfahren zur objektiven Bestimmung des Anfallsendpunkts. Vor diesem Hintergrund ergeben sich folgende zentrale Forschungsfragen:

1) In welchem Ausmaß lässt sich der Anfallsendpunkt mithilfe eines auf maschinellem Lernen basierenden Verfahrens zuverlässig bestimmen, und treten signifikante Abweichungen gegenüber der visuellen Beurteilung erfahrener klinischer Experten sowie den Ausgaben etablierter gerätebasierter Systeme auf?
2) Inwieweit lassen sich etablierte Qualitätsparameter wie Krampfdauer, postiktaler Suppressionsindex sowie amplituden- und energiebasierte Indizes durch digitale Biosignalanalyse zuverlässig bestimmen, und bestehen signifikante Unterschiede zu den Ergebnissen manueller Bewertungen oder geräteinterner Berechnungen?
3) Welche methodischen Erweiterungen ermöglicht das entwickelte Framework für die Identifikation und Validierung zusätzlicher Qualitätsparameter sowie für die Weiterentwicklung prädiktiver klinischer Entscheidungshilfen im Kontext der EKT?

1.1.6 Zielsetzung

Ziel dieser Arbeit ist es, ein digitales Framework zu konzipieren und zu implementieren, welches EKT induzierte Krampfanfälle auf der Grundlage von EEG-Aufzeichnungen automatisch erkennt und einen einheitlich definierten Anfallsendpunkt bestimmt. Darauf aufbauend sollen etablierte Qualitätsparameter wie Krampfdauer, postiktaler Suppressionsindex sowie energie- und amplitudenbasierte Indizes und kohärenzbasierte Maße objektiv und reproduzierbar berechnet werden. Die mit diesem Analyse-Framework gewonnenen Kennwerte werden mit den Referenzangaben des Stimulationsgeräts und der klinischen Beurteilung erfahrener EKT-Experten verglichen, um die Zuverlässigkeit des Verfahrens und die Übereinstimmung mit der bisherigen Praxis zu bewerten. Darüber hinaus soll das Framework als Grundlage für zukünftige methodische Erweiterungen und prädiktive Entscheidungsmodelle dienen.

1.2 Material und Methoden

Die zugrundeliegende Originalpublikation (Kayser et al., 2025) basiert auf einer retrospektiven Analyse von EEG-Daten sowie behandlungsbezogenen Parametern, die im Rahmen von 1.634 EKT-Behandlungen bei 116 stationären Patienten erhoben wurden. Die Datenerfassung erfolgte an zwei Behandlungszentren: dem Universitätsklinikum Hamburg-Eppendorf (UKE) und dem Universitätsklinikum Bonn (UKB).

Die methodische Umsetzung stützt sich auf einem mehrstufigen Framework, der von der standardisierten EKT-Datenerhebung über die Segmentierung der EEG-Aufzeichnungen und die automatisierte Anfallserkennung mit Verfahren des maschinellen Lernens bis zur Berechnung etablierter Qualitätsparameter reicht. Abbildung 1 zeigt die Abfolge der Verarbeitungsschritte und Analyse von der Datenerhebung bis zur Bewertung der Krampfqualität in zusammenfassender Form.

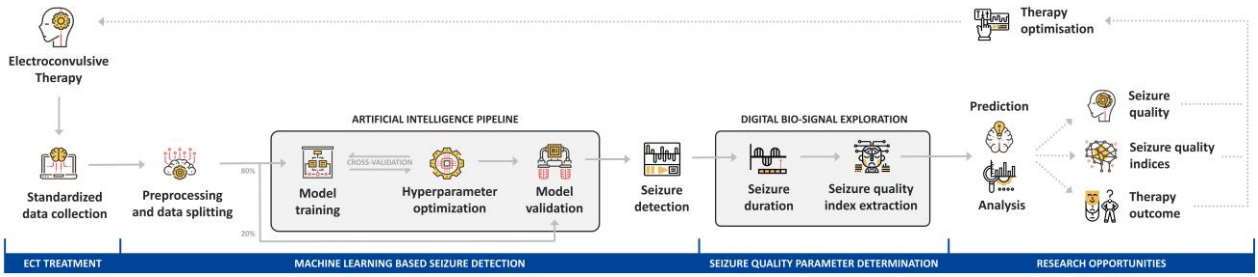

Abb. 1: Schematischer Gesamtworkflow des digitalen Analyse-Frameworks von der Datenerhebung bis zur automatisierten Krampferkennung und Auswertung der EKT-Qualitätsparameter (nach Kayser et al., 2025).

Im dargestellten Ablauf werden zunächst die, während der EKT standardisiert erhobenen Daten vorverarbeitet, in Trainings- und Testdaten aufgeteilt und durch Merkmalsextraktion in eine strukturierte Form überführt. Daran schließt sich die ML-Pipeline innerhalb des Analyse-Frameworks mit Modelltraining, Hyperparameteroptimierung und Validierung der Klassifikationsmodelle an. Auf dieser Grundlage erfolgt die automatisierte Krampferkennung mit der Berechnung von Krampfdauer und weiteren Krampfqualitätsparametern sowie deren statistischer Auswertung. Die resultierenden Modelle und Kennwerte bilden

eine Grundlage für weiterführende Forschungsarbeiten und eröffnen Perspektiven für einen zukünftigen klinischen Einsatz des digitalen Frameworks, mit dem Ziel, die EKT weiter zu optimieren und stärker zu individualisieren.

Das Framework wurde modular realisiert. In den folgenden Abschnitten werden die zentralen methodischen Schritte erläutert.

1.2.1 Ethik und regulatorische Rahmenbedingungen

Für diese retrospektive Studie mit anonymisierten Patientendaten sah die zuständige Ethikkommission von der Einholung individueller Einwilligungen und eines gesonderten Ethikvotums ab. Datenerhebung und -auswertung standen im Einklang mit den geltenden gesetzlichen und institutionellen Bestimmungen. Die Behandlung erfolgte nach hausinternen Standard Operating Procedures zur EKT, die in den beteiligten Kliniken etabliert sind, und unter Berücksichtigung der einschlägigen S3 Leitlinien (Gaebel et al., 2019; Schneider et al., 2017).

1.2.2 Elektrokonvulsionstherapie

1.2.2.1 Patientenkollektiv und Indikation

Die Studienkohorte umfasste Patienten mit therapieresistenter Depression nach ICD 10 mit den Diagnosen F32.2 bis F32.3 und F33.1 bis F33.4 sowie Patienten mit bipolaren Störungen und schweren depressiven Episoden nach ICD 10 mit den Diagnosen F31.4 und F31.8. Alle eingeschlossenen Patienten erhielten zwischen 2019 und 2022 an den Studienzentren UKB und UKE eine Elektrokonvulsionstherapie Serie. Die Indikationsstellung erfolgte durch einen in der EKT erfahrenen Facharzt für Psychiatrie auf Grundlage einer individualisierten klinischen Einschätzung unter Berücksichtigung evidenzbasierter Empfehlungen und etablierter Versorgungsleitlinien (BÄK et al., 2022). Die Aufklärung zur EKT sowie zur Narkose wurde durch die behandelnden Psychiater bzw. den Anästhesisten durchgeführt und schriftlich dokumentiert. Bei fehlender Einwilligungsfähigkeit wurde die Entscheidung im Rahmen der gesetzlichen Betreuung getroffen. In die Analyse wur-

den ausschließlich Index EKT Serien einbezogen. Erhaltungsbehandlungen wurden ausgeschlossen. Jede EKT-Serie bestand aus 5 bis 29 Sitzungen. Das Behandlungsintervall betrug am UKB zwei und am UKE drei Sitzungen pro Woche. Überschritt der Abstand zwischen zwei Sitzungen zehn Tage und lag keine anderslautende Dokumentation vor, wurde die Serie als beendet gewertet. Die betreffende Sitzung und alle nachfolgenden Behandlungen wurden in diesem Fall aus dem Datensatz entfernt. Wurden innerhalb einer Sitzung mehrere Stimulationsdurchgänge durchgeführt, ging ausschließlich die letzte Stimulation in die Auswertung ein.

1.2.2.2 Technische Durchführung und Stimulationsparameter

Als Stimulationsgerät kam in beiden Studienzentren das Thymatron® System IV (Somatics LLC, Lake Bluff, Illinois, USA) zum Einsatz. Die Elektroden wurden je nach Indikation bitemporal, bifrontal, unilateral rechts oder links oder in einer links-anterior-rechts-temporalen Konfiguration platziert (Sackeim et al., 2000; Swartz, 2009). Die Stimulation erfolgte mit einem Rechtecksignal (Squire und Zouzounis, 1986) und einer kurzen Impulsdauer von mindestens 0,25 ms (Niemantsverdriet et al., 2011). Die Stimulationsfrequenz lag zwischen 10 und 70 Hz. Die Stimulationsstärke variierte zwischen 5 % und 100 % der Maximalladung von 504 mC und erreichte in Einzelfällen bis zu 200 %. Wurde kein oder nur ein unzureichender Krampfanfall ausgelöst, konnte eine erneute Stimulation mit einer um etwa 10 bis 20 % erhöhten Energie erfolgen. Aufgrund der Refraktärität des Nervengewebes wurden Wiederholungen frühestens nach einem zeitlichen Abstand von mindestens einer Minute durchgeführt (Kranaster et al., 2012).

Die Krampfschwelle wurde zu Beginn der Serie individuell bestimmt (Aten et al., 2015). Am UKB erfolgte die Bestimmung mittels Dosistitration, am UKE anhand der Altersmethode. Bei der Titrationsmethode begann die Stimulationsdosis bei 5 % und wurde, sofern erforderlich, schrittweise auf 10 % und 20 % gesteigert, bis ein suffizienter Krampfanfall ausgelöst wurde (Beale et al., 1994). Trat bei 10 % kein Krampf auf, jedoch bei 20 % eine ausreichende Reaktion, wurde die Krampfschwelle rechnerisch auf 15 % festgelegt. Innerhalb einer Therapiesitzung waren maximal drei Stimulationsversuche zulässig. In den

nachfolgenden Sitzungen wurden die Stimulationen bei unilateraler Platzierung mit mindestens dem Vierfachen und bei bitemporaler Stimulation mit dem Eineinhalb- bis Zweifachen der ermittelten Krampfschwelle durchgeführt. Bei der Altersmethode entsprach die initiale Stimulationsdosis in Prozent dem Lebensalter des Patienten. Bei bitemporaler Elektrodenplatzierung wurde dieser Wert halbiert (Swartz und Michael, 2013).

1.2.2.3 Anästhesie und Monitoring

Die Behandlung erfolgte nach ausreichender Präoxygenierung unter Allgemeinanästhesie (Kellner et al., 2009; Pape und Kittel-Schneider, 2023). Hierfür wurde entweder Propofol (1 - 2 mg/kg Körpergewicht i. v.) oder eine 1 zu 1 Kombination aus Ketamin und Propofol in gleicher Dosierung eingesetzt verabreicht (Bailine et al., 2003; Yalcin et al., 2012). Zur Muskelrelaxation kam Succinylcholin (1 mg/kg Körpergewicht i. v.) oder Rocuroniumbromid (0,5 - 1 mg/kg Körpergewicht i. v.) zur Anwendung (Grundmann und Oest, 2007). Optional wurde Glycopyrroniumbromid (0,2 mg i. v.) verabreicht, um eine übermäßige Speichelproduktion zu vermindern und das Aspirationsrisiko zu reduzieren. Nach der Induktion erfolgte eine Maskenbeatmung mit 100 % Sauerstoff, um eine ausreichende Oxygenierung bis zum Wirkeintritt des Muskelrelaxans sicherzustellen. Nach vollständiger Relaxierung wurden ein Zahnschutz eingelegt, die Impedanz am Stimulationsgerät überprüft und die Stimulation gemäß den festgelegten Parametern ausgelöst. Unmittelbar nach Krampfende wurde erneut maskenbeatmet, um die Sauerstoffversorgung während der postiktalen Phase bis zur Ausleitung der Narkose aufrechtzuerhalten. Sofern ein nicht depolarisierendes Muskelrelaxans verwendet wurde und eine neuromuskuläre Restblockade bestand, erfolgte unter relaxometrischer Kontrolle eine Antagonisierung mit Sugammadex (2 - 4 mg/kg Körpergewicht i. v.). Verlängerte Krampfanfälle und Status epileptici wurden medikamentös behandelt (Aloysi et al., 2014). Nach Abschluss der Behandlung erfolgten die Narkoseausleitung und eine postanästhesiologische Nachbetreuung. Während der gesamten Behandlung erfolgte eine kontinuierliche Überwachung des Herz-Kreislauf-Systems mittels Elektrokardiogramm, nicht-invasiver Blutdruckmessung und peripherer Sauerstoffsättigung sowie eine Überwachung der Hirnaktivität über ein zweikanaliges EEG.

1.2.2.4 Klinische Verlaufsdokumentation

Vor jeder EKT-Sitzung wurde der Clinical Global Impression Severity Score zur unmittelbaren klinischen Beurteilung des Schweregrads der Erkrankung und zur longitudinalen Dokumentation des Behandlungsverlaufs erhoben (CGI-S; Busner und Targum, 2007). Der CGI-S basiert auf einer siebenstufigen Likert Skala, auf der 1 keine klinisch relevante Symptomatik und 7 eine maximal ausgeprägte Erkrankung abbildet. Dies ermöglicht eine standardisierte Verlaufsbewertung über den gesamten Therapiezeitraum. Für die vorliegende Auswertung wurden Patienten als Responder definiert, wenn sie eine Verbesserung um mindestens zwei CGI-S-Punkte gegenüber dem Ausgangswert zeigten oder bei der letzten Sitzung einen Wert von 3, entsprechend einer leichten Symptomatik, oder weniger erreichten (Morrens et al., 2022).

1.2.3 Datenerfassung und EEG-Aufzeichnung

Die Erfassung sämtlicher behandlungs- und patientenbezogener Daten erfolgte mithilfe von GENET-GPD, einer speziell für die Elektrokonvulsionstherapie entwickelten Datenbank und Dokumentationsplattform zur standardisierten Erhebung relevanter Parameter (Freundlieb et al., 2023; Kiebs und Freundlieb, 2021). Neben Stimulationsparametern wie Elektrodenposition, Stimulationsenergie, Stromstärke und Pulsbreite wurden Vitalzeichen, soziodemografische Variablen sowie klinische Scores, darunter der CGI-S (siehe Abschnitt 1.2.2.4), erfasst. Die vom Stimulationsgerät mit einer Abtastrate von 200 Hz aufgezeichneten EEG-Signale wurden über eine serielle Schnittstelle in GENET-GPD übertragen und umfassten eine zweikanalige fronto-mastoidale Ableitung. Zur Reduktion von Artefakten und zur Sicherung der Signalqualität wurden ein 50 Hz Notch Filter zur Unterdrückung von Netzbrummen und ein 30 Hz Tiefpassfilter zur Entfernung hochfrequenter Anteile eingesetzt. Zusätzlich zu den Rohdaten berechnete das Stimulationsgerät mehrere Anfallsqualitätskennwerte, darunter die Krampfdauer (siehe auch Abschnitt 1.2.6.1) und den Postiktaler Suppressionsindex (PSI, siehe auch Abschnitt 1.2.6.2).

Zur Validierung der geräteinternen Berechnungen wurden der postiktale Suppressionsindex sowie der Anfallsendpunkt zusätzlich durch eine verblindete visuelle Beurteilung erfasst. Ein erfahrener Rater schätzte den Grad der EEG-Suppression in Schritten von 10

% von 0 % ohne Suppression bis 100 % vollständige Suppression ein. Die Krampfdauer wurde visuell in Sekunden bestimmt. Der Rater legte das Anfallsende anhand des Abklingens hochfrequenter Spike Wave Komplexe und des Übergangs zu flacher, niederamplitudiger EEG-Aktivität fest. Dieser Übergang war typischerweise durch eine abrupte Reduktion von Frequenz und Amplitude oder das Auftreten einer nahezu isoelektrischen Nulllinie gekennzeichnet, abhängig vom individuellen EEG-Muster des Patienten, und wurde als Übergang von hochfrequenter EEG-Aktivität zur postiktalen Suppression definiert. Er markierte zugleich das Ende der iktalen Phase.

1.2.4 Datenaufbereitung

Die für diese Studie erhobenen Daten wurden in einer wissenschaftlichen Python Umgebung der Version 3.7.6 auf dem Linux basierten High Performance Computing Cluster BONNA der Universität Bonn aufbereitet und analysiert (Rossum und Drake, 2010). Sämtliche Verarbeitungsschritte, einschließlich statistischer Auswertungen, Modellierungen und numerischer Berechnungen, wurden auf diesem System durchgeführt. Rechenintensive Prozesse wurden unter anderem mithilfe der Bibliothek `Joblib` parallelisiert, um die Rechenzeiten zu verkürzen (Joblib Development Team Joblib Development Team, 2020). Für die Verarbeitung der Datensätze wurden insbesondere `pandas` und `NumPy` eingesetzt (Harris et al., 2020; McKinney, 2010; Reback et al., 2021). Die EEG-Signale, welche im European Data Format vorlagen, wurden mithilfe von `pyEDFlib` importiert (Kemp et al., 1992; Kemp und Olivan, 2003; Nahrstaedt et al., 2020). Zur Sicherstellung von Reproduzierbarkeit und Vergleichbarkeit aller Verarbeitungs-, Modellierungs- und Auswertungsschritte wurde ein zufälliger Startwert (engl. *random seed*) gesetzt, der für sämtliche Berechnungen einheitlich verwendet wurde.

Fehlende Werte, etwa für Alter, Geschlecht, Diagnose oder Elektrodenplatzierung, wurden mithilfe von Last Observation Carried Forward und Next Observation Carried Backward imputiert (Little und Rubin, 2020). Bei unvollständigen Werten der CGI-S kam hingegen eine lineare Interpolation zur Anwendung.

Für die Krampferkennung und die Berechnung der Krampfparameter wurde primär ein Einkanal-Ansatz verwendet. Dabei wurde das EEG Signal der rechten Hemisphäre genutzt, da bei überwiegend rechts unilateraler Stimulation dort stärkere iktale und postiktale EEG Veränderungen auftreten (Nobler et al., 1993; Perera et al., 2004). Verfahren zur Kohärenzanalyse bildeten eine Ausnahme, da hierfür die EEG-Signale beider Ableitungen berücksichtigt wurden (siehe Abschnitt 1.2.6.8).

Aufzeichnungen mit einer Gesamtdauer von über 120 Sekunden, wurden ausgeschlossen, da sie als verlängert gelten und mit kognitiven Beeinträchtigungen assoziiert werden (Aloysi et al., 2014; Freeman, 1995; Grözinger et al., 2013; Rasmussen, 2002). Damit gingen atypische Krampfverläufe nicht in die Analyse ein.

1.2.5 Anfallserkennung

Zur Detektion der induzierten Anfälle wurden die aufgezeichneten EEG-Daten zunächst in gleich große, nicht überlappende Segmente von jeweils 1,28Sekunden Dauer unterteilt. Diese Segmentlänge bietet eine ausreichend feine zeitliche Auflösung, um iktale EEG-Ereignisse abzubilden, und stellt zugleich genügend Datenpunkte für statistische und frequenzanalytische Auswertungen zur Verfügung.

Die anschließende Segmentzuordnung stützte sich auf die vom Stimulationsgerät ermittelten Anfallsendpunkte. Alle EEG-Segmente vor diesem Endpunkt wurden mit dem Label „iktal“ versehen, da sie direkte Anfallsaktivität oder zeitlich angrenzende Abschnitte der iktalen Phase enthielten. Segmente nach dem dokumentierten Anfallsende wurden als „nicht-iktal“ klassifiziert, da sie postiktale Aktivität oder anfallsfreie Ruhephasen abbildeten. Das erste Segment jeder Aufzeichnung erhielt zusätzlich das Label „stimulation“, da es den Zeitraum der aktiven Stimulation umfasste. Aufzeichnungen ohne eindeutig bestimmbaren Anfallsendpunkt wurden der Kategorie „nicht-spezifiziert“ zugeordnet. Die Label „stimulation“ und „nicht-spezifiziert“ wurden nicht in die binäre Klassifikation einbezogen und aus dem weiteren Datensatz entfernt.

1.2.5.1 Maschinelles Lernen

Zur binären Klassifikation der EEG-Segmente wurde ein Verfahren aus dem Bereich des überwachten maschinellen Lernens eingesetzt. Unter maschinellem Lernen werden Methoden verstanden, bei denen Modelle aus Trainingsdaten eigenständig Muster und Zusammenhänge ableiten, ohne dass für jede mögliche Konstellation explizite Regeln vorgegeben werden. Beim überwachten Lernen (engl. *supervised learning*) erfolgt das Training anhand eines Datensatzes mit bekannten Klassenzugehörigkeiten (engl. *ground truth*). Im vorliegenden Fall lagen die EEG-Segmente mit den Labels „iktal“ und „nicht-iktal“ wie in Abschnitt 1.2.5 beschrieben vor. Die Modelle lernten aus diesen Beispielen Merkmalskonstellationen zu identifizieren, die typischerweise mit iktalen oder nicht-iktalen Phasen einhergehen, und nutzten dieses Wissen zur Klassifikation bislang ungesehener Segmente. Die EEG-Segmente wurden nicht direkt über die Rohsignale verarbeitet, sondern mithilfe einer strukturierten Merkmalsextraktion beschrieben, sodass jedes Segment als Punkt in einem mehrdimensionalen Merkmalsraum abgebildet wurde.

1.2.5.2 Merkmalsextraktion

Zur Merkmalsextraktion wurde die `TSFEL` Bibliothek verwendet (Barandas et al., 2020). Sie ermöglicht die automatische Berechnung zahlreicher spektraler, statistischer und zeitbasierter Kennwerte aus EEG-Segmenten. Diese so gewonnenen Merkmale (engl. *features*) bilden unterschiedliche Aspekte der neuronalen Aktivität ab und schaffen die Grundlage für eine algorithmische Unterscheidung zwischen iktalen und nicht-iktalen Phasen.

Die extrahierten Merkmale lassen sich in vier Kategorien unterteilen:

- Energiebasierte Merkmale:
  Diese Merkmale beschreiben die Energieverteilung innerhalb eines Segments. Hierzu zählen die absolute Energie (engl. *absolute energy*), die Gesamtenergie (engl. *total energy*), die Fläche unter der Kurve (engl. *area under the curve*; AUC) und der Effektivwert (engl. *root mean square*) als Maß für die mittlere Effektivamplitude. Ergänzend wurden Minimum und Maximum der Amplitude sowie Kennwerte berechnet, die die

maximale Spektralleistung (engl. *max power spectrum*) und die Bandbreite der spektralen Leistung charakterisieren (engl. *power bandwidth*).

- Statistische Merkmale:
  Diese Merkmalsgruppe beschreibt grundlegende statistische Eigenschaften der Signalverteilung. Dazu gehören Lage- und Streuungsmaße wie der Mittelwert (engl. *mean*), der Median (engl. *median*), die Standardabweichung (engl. *standard deviation*) und die Varianz (engl. *variance*). Die Form der Verteilung wird durch die Schiefe (engl. *skewness*) und die Wölbung (engl. *kurtosis*) beschrieben. Zusätzlich wurden differenzbasierte Merkmale extrahiert, die Veränderungen innerhalb des Segments quantifizieren. Dazu zählen die mittlere absolute Abweichung (engl. *mean absolute deviation*), die mediane absolute Abweichung (engl. *median absolute deviation*) sowie die Summe der absoluten Differenzen (engl. *sum absolute difference*). Die mittlere absolute Differenz zwischen benachbarten Werten (engl. *mean absolute difference*) und der Median der Differenzen (engl. *median difference*) erfassen direkt die Variabilität aufeinanderfolgender Werte. Die Signaldistanz (engl. *signal distance*) bezeichnet die Summe der absoluten Differenzen zwischen jeweils benachbarten Signalwerten innerhalb eines Segments.
- Frequenzbasierte Merkmale:
  Frequenzmerkmale analysieren die spektrale Zusammensetzung des Signals. Dazu gehören die Fundamentalfrequenz (engl. *fundamental frequency*), die maximale Frequenz (engl. *maximum frequency*) und die mediane Frequenz (engl. *median frequency*), die jeweils die dominierenden Frequenzkomponenten erfassen. Zusätzlich wurde der Spektralschwerpunkt (engl. *centroid*) bestimmt, der die mittlere Lage der spektralen Energie angibt. Ein weiteres Merkmal dieser Gruppe ist die Nulldurchgangsrate (engl. *zero crossing rate*), die die Häufigkeit der Vorzeichenwechsel innerhalb des Segments quantifiziert. Zur Charakterisierung der Signalkomplexität wurde darüber hinaus die Entropie (engl. *entropy*) berechnet, die die Unregelmäßigkeit der Amplitudenverteilung innerhalb eines Segments abbildet.
- Temporale Strukturmerkmale:
  Diese Merkmale beschreiben die innere Struktur und die zeitliche Dynamik innerhalb eines Segments. Die Autokorrelation (engl. *autocorrelation*) misst die Ähnlichkeit des

Signals mit sich selbst über verschiedene Zeitabstände hinweg und liefert so Hinweise auf periodische Muster. Ergänzend wurden die Anzahl der positiven Umkehrpunkte (engl. *positive turning points*) und negativen Umkehrpunkte (engl. *negative turning points*) bestimmt, die die Häufigkeit von Richtungswechseln anzeigen. Die benachbarten Signalspitzen (engl. *neighbourhood peaks*) bezeichnen die Anzahl von lokalen Maxima, die in unmittelbarer zeitlicher Nähe zueinander auftreten. Die Spitzen-zu-Spitzen-Distanz (engl. *peak to peak distance*) beschreibt den Abstand zwischen zwei aufeinanderfolgenden lokalen Maxima. Die Steigung (engl. *slope*) schließlich quantifiziert die mittlere Änderungsrate innerhalb des Zeitfensters.

Eine detaillierte Liste der extrahierten Merkmale und deren Definitionen findet sich in Tabelle S2 des Supplements der Originalpublikation (Kayser et al., 2025). Die so gewonnenen Merkmale wurden zu einer Merkmalsmatrix zusammengeführt, in der jede Zeile einen Merkmalsvektor und damit ein einzelnes EEG-Segment repräsentiert und jede Spalte ein einzelnes Merkmal abbildet. Jedem Segment war zusätzlich ein Klassenlabel zugeordnet, das angab, ob es der Klasse „iktal“ oder „nicht-iktal“ angehörte, wie in Abschnitt 1.2.5 beschrieben.

Im Rahmen der initialen Merkmalsselektion wurden die extrahierten Merkmale anhand statistischer Kenngrößen wie Minimum, Maximum, Mittelwert, Standardabweichung und Varianz beschrieben, die in Supplement B.1 der Originalpublikation zusammengefasst sind (Kayser et al., 2025). Zur Reduktion potenziell irrelevanter Information wurden Merkmale mit geringer Varianz von höchstens 20 % ausgeschlossen. Stark korrelierte Merkmale mit einem Korrelationskoeffizienten von mindestens 0,95 wurden ebenfalls entfernt, um Redundanz zu verringern und die Dimensionalität des Merkmalsraums zu begrenzen.

1.2.5.3 Balancierung und Aufteilung der Daten

Zur Herstellung einer ausgewogenen Klassenverteilung und Reduktion potenzieller Verzerrungen wurde eine randomisierte Unterabtastung (engl. *random undersampling*) angewendet (Mohammed et al., 2020; Saripuddin et al., 2021). Dabei wurden Stichproben der überrepräsentierten Klasse entfernt, bis beide Klassen gleich häufig vertreten waren. Die Umsetzung erfolgte mit der Python-Bibliothek `imbalanced-learn` (Lemaitre et al.,

2016). Anschließend wurde die balancierte Merkmalsmatrix im Verhältnis 80:20 in Trainings- und Testdaten aufgeteilt. Die Merkmalsvektoren sowie ihre zugehörigen Labels wurden zufällig auf Trainings- und Testsets verteilt, unabhängig von der zeitlichen Abfolge oder der Position innerhalb der ursprünglichen EEG-Aufzeichnung. Die Aufteilung erfolgte segmentbasiert, wobei die einzelnen Segmente als analytisch unabhängig behandelt wurden, um potenziellen Verzerrungen durch temporale Autokorrelation innerhalb zusammenhängender EEG-Sequenzen entgegenzuwirken.

1.2.5.4 Aufbau des digitalen Analyse-Frameworks

Alle Verarbeitungsschritte wurden in einem standardisierten Analyse-Framework gebündelt, das eine reproduzierbare und transparente Datenverarbeitung ermöglicht (siehe Abbildung 1).

1.2.5.5 Datenvorverarbeitung

Die Vorverarbeitung der Daten (engl. *preprocessing*) innerhalb des Frameworks umfasste eine lineare Normalisierung der Merkmalsmatrix und eine anschließende Hauptkomponentenanalyse (engl. *Principal Component Analysis*). Im ersten Schritt wurden alle Merkmale auf einen Wertebereich zwischen 0 und 1 skaliert, um Verzerrungen durch unterschiedliche Skalierungen der Merkmale zu vermeiden (Maharana et al., 2022). Anschließend erfolgte eine Hauptkomponentenanalyse (Pearson, 1901; Rokhlin et al., 2008), wobei die Anzahl der Komponenten datengetrieben so gewählt wurde, dass mindestens 80 % der Gesamtvarianz abgedeckt waren. Diese Schwelle diente der gezielten Reduktion der Dimensionalität bei gleichzeitigem Erhalt zentraler Informationsanteile.

1.2.5.6 Algorithmen des maschinellen Lernens

Für die Klassifikationsaufgabe wurden fünf verschiedene Algorithmen des maschinellen Lernens aus der `scikit-learn` Bibliothek herangezogen (Pedregosa et al., 2011):

- Der Entscheidungsbaum (engl. *Decision Tree*; DT) segmentiert den Merkmalsraum rekursiv anhand der Eingangsmerkmale und erzeugt eine baumartige Struktur, in der jeder Pfad vom Wurzelknoten zu einem Blatt einer Entscheidungsregel entspricht (Breiman et al., 2017; Loh, 2011; Quinlan, 1986). Entscheidungsbäume sind gut interpretierbar, neigen jedoch bei großer Tiefe zu Überanpassung und dienen in dieser Arbeit vor allem als einfaches Referenzmodell.
- Der Random Forest basiert auf einem Ensemble vieler Entscheidungsbäume, die auf zufällig gezogenen Teilmengen der Trainingsdaten und Merkmale trainiert werden (Breiman, 2001; Liaw und Wiener, 2002). Die finale Vorhersage ergibt sich aus der Mehrheitsentscheidung aller Einzelbäume. Durch diese Aggregation werden Varianz und Überanpassung einzelner Bäume reduziert, sodass sich das Verfahren besonders für komplexe und hochdimensionale Datensätze eignet (Cutler et al., 2007).
- Der Stützvektor Klassifikator (engl. *Support Vector Classifier*; SVC) nutzt das Prinzip der maximalen Margin Trennung und bestimmt eine Entscheidungsgrenze, die die Klassen mit größtmöglichem Abstand zu den nächstgelegenen Trainingspunkten trennt (Cortes und Vapnik, 1995; Vapnik, 1998). Durch den Einsatz von Kernelfunktionen können auch nicht lineare Zusammenhänge erfasst werden, indem die Daten in einen höherdimensionalen Raum abgebildet werden, in dem eine lineare Trennung möglich ist (Hastie et al., 2001).
- Der k-Nächste-Nachbarn-Klassifikator (engl. *k-Nearest-Neighbors*; KNN) ordnet einem neuen Datenpunkt die Klasse zu, die unter seinen k nächstgelegenen Nachbarn im Merkmalsraum am häufigsten vertreten ist (Cover und Hart, 1967; Omohundro, 1989). Das Verfahren bildet kein explizites Modell, sondern nutzt die Trainingsdaten direkt als Referenz. Es kann komplexe, nicht lineare Entscheidungsgrenzen abbilden, reagiert jedoch empfindlich auf die Wahl des Parameters k und auf Skalierungseffekte in den Merkmalen (Duda et al., 2001).
- Der Gradient Boosting Klassifikator (engl. *Gradient Boosting Classifier*; GBC) optimiert die Modellleistung durch ein iteratives Verfahren, bei dem viele schwache Entscheidungsbäume sequenziell trainiert und zu einem starken Gesamtmodell kombiniert werden (Friedman, 2001; Mason et al., 1999; Natekin und Knoll, 2013). Jeder neue Baum fokussiert auf die Residualfehler der bisherigen Vorhersagen, sodass schwer vorhersagbare Instanzen schrittweise besser abgebildet werden. Die Lernrate steuert

den Grad dieser Anpassung und erlaubt eine Balance zwischen Anpassungsfähigkeit und Generalisierungsfähigkeit.

Die standardisierte Einbindung sämtlicher Klassifikatoren in das Framework ermöglichte eine konsistente Trainings- und Bewertungsgrundlage für den anschließenden Vergleich ihrer Klassifikationsleistung.

1.2.5.7 Training der Klassifikationsmodelle

Auf Grundlage der Trainingsdaten wurde für jeden Algorithmus ein Klassifikationsmodell trainiert, das die Zuordnung der Segmente zu den Klassen „iktal“ und „nicht-iktal“ ermöglichen sollte. Um eine geeignete Modellkonfiguration zu bestimmen und das Risiko von Überanpassung zu reduzieren, wurde für jedes Verfahren eine Hyperparameteroptimierung mittels einer fünffachen Kreuzvalidierung in Kombination mit einer Grid-Search durchgeführt (Arlot und Celisse, 2010; Geisser, 1975; Pedregosa et al., 2011). Hierbei wurde das Trainingsset in fünf gleich große Teilmengen (engl. *folds*) unterteilt. Die Modelle wurden in jeder Iteration auf vier Folds trainiert und auf dem verbleibenden Fold validiert, bis jede Teilmenge einmal als Validierungsdatensatz fungierte (Wilimitis und Walsh, 2023). Im Rahmen der Grid-Search wurden alle Kombinationen vordefinierter Hyperparameter evaluiert; die geprüften Parameterbereiche sind in Tabelle 2 aufgeführt. Zur Auswahl der optimalen Hyperparameter wurde die Modellgüte anhand des Matthews Korrelationskoeffizienten bewertet, der in nachfolgenden Abschnitt näher beschrieben ist (Chicco und Jurman, 2020).

1.2.5.8 Evaluation und Leistungsbewertung der Modelle

Nach Abschluss der Optimierung wurden die trainierten Klassifikationsmodelle auf ein unabhängiges Testset angewendet, um deren finale prädiktive Leistung außerhalb der Trainings- und Validierungsdaten zu bestimmen. Zur Bewertung der Modellgüte diente die Konfusionsmatrix, die die Anzahl korrekt und inkorrekt klassifizierter Segmente in beiden Klassen, „iktal“ und „nicht-iktal“, abbildet (Kubat, 2015; Powers, 2020). Sie umfasst vier Elemente. True Positives (TP) bezeichnen korrekt als „iktal“ klassifizierte Segmente. False

Positives (FP) sind fälschlicherweise als „iktal“ klassifizierte Segmente, die tatsächlich nicht-iktal sind. False Negatives (FN) entsprechen fälschlicherweise als „nicht-iktal“ klassifizierten Segmenten, die tatsächlich iktal sind. True Negatives (TN) bezeichnen korrekt als „nicht-iktal“ klassifizierte Segmente.

Diese Einteilung bildet die Grundlage für verschiedene ML-Metriken, die für eine umfassende Leistungsbewertung der Modelle herangezogen wurden (Sokolova und Lapalme, 2009):

- Die Accuracy beschreibt den Anteil der insgesamt korrekt klassifizierten Segmente an der Gesamtzahl der Segmente. Sie wird berechnet als:

  $$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

  Der Wertebereich liegt zwischen 0 und 1. Ein Wert von 1 entspricht einer perfekten Klassifikation, ein Wert von 0 einer vollständig fehlerhaften Klassifikation. Bei zufälliger Klassifikation ohne Trennschärfe ergibt sich bei ausgewogener Klassenverteilung typischerweise ein Wert um 0,5. Die Accuracy eignet sich vor allem bei ausgewogenen Klassenverteilungen. Bei deutlich unausgeglichenen Klassen kann sie hingegen irreführend sein, da ein Modell allein durch die bevorzugte Zuordnung zur häufigeren Klasse eine hohe Accuracy erreichen kann, ohne eine klinisch sinnvolle Trennleistung zu erbringen.

- Die Precision, auch positiver prädiktiver Wert (engl. *Positive Predictive Value*; PPV) genannt, gibt an, welcher Anteil der als iktal klassifizierten Segmente tatsächlich iktal ist. Sie misst die Genauigkeit positiver Vorhersagen und wird berechnet als:

  $$Precision = PPV = \frac{TP}{TP + FP}$$

  Der Wertebereich reicht ebenfalls von 0 bis 1. Eine hohe Precision bedeutet, dass nur wenige nicht-iktale Segmente fälschlicherweise als iktal klassifiziert werden. Eine niedrige Precision weist auf viele falsch positive Vorhersagen hin. Precision ist insbesondere dann von Bedeutung, wenn falsch positive Fehler begrenzt werden sollen.

- Die Sensitivität, auch Recall oder Richtig-Positiv-Rate (engl. *True Positive Rate*; TPR) beschreibt den Anteil der tatsächlich iktalen Segmente, die korrekt als iktal erkannt werden. Sie ergibt sich aus:

$$Sensitivität = \ Recall = \mathrm{TPR} = \frac{TP}{TP + FN}$$

Auch hier liegt der Wertebereich zwischen 0 und 1. Eine hohe Sensitivität bedeutet, dass nur wenige Anfälle übersehen werden, eine niedrige Sensitivität dagegen, dass ein relevanter Teil der tatsächlichen Anfälle nicht erkannt wird.

- Der F1-Score kombiniert Precision und Sensitivität in einer einzigen Metrik, indem er deren harmonisches Mittel berechnet (Sasaki, 2007). Dies ermöglicht eine ausgewogene Bewertung der Modellleistung, insbesondere wenn ein Kompromiss zwischen Präzision und Sensitivität erforderlich ist (Lipton et al., 2014). Die Berechnung erfolgt nach:

$$F_1 = \frac{2 \times Precision \times Sensitivität}{Precision + Sensitivität} = \frac{2 \times TP}{(2 \times TP) + FP + FN}$$

Der Wertebereich reicht von 0 bis 1. Ein hoher F1-Score zeigt, dass das Modell sowohl eine gute Precision als auch einen hohen Recall erreicht. Ist der Wert niedrig, bedeutet dies, dass entweder die False-Positive-Rate hoch ist (niedrige Precision) oder dass viele Anfälle übersehen werden (niedriger Recall). Der F1-Score ist besonders hilfreich, wenn die Klassenverteilung unausgewogen ist, da er sowohl die Fähigkeit zur korrekten Identifikation positiver Fälle als auch zur Vermeidung falsch-positiver Vorhersagen berücksichtigt.

- Der Matthews Korrelationskoeffizient (engl. *Matthews Correlation Coefficient*; MCC) ist eine umfassende Metrik zur Gesamtbewertung eines Klassifikationsmodells (Matthews, 1975), da er alle vier Einträge der Konfusionsmatrix berücksichtigt (Chicco und Jurman, 2020). MCC ist besonders nützlich für unausgewogene Klassenverteilungen, da er – anders als Accuracy – nicht durch eine dominierende Klasse verzerrt wird. Die Berechnung erfolgt mit:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Der Wertebereich liegt zwischen -1 und +1, wobei ein Wert von +1 eine perfekte Klassifikation, ein Wert von 0 eine zufällige Klassifikation und ein Wert von -1 eine vollständig fehlerhafte Klassifikation bedeutet. MCC ist besonders aussagekräftig, wenn die Daten stark unausgeglichen sind, da er sicherstellt, dass sowohl richtige als auch falsche Klassifikationen in beiden Klassen gleichmäßig berücksichtigt werden.

- Die Grenzwertoptimierungskurve (engl. *Receiver Operating Characteristic*; ROC) beschreibt die Beziehung zwischen der Richtig-Positiv-Rate (engl. *True Positive Rate*; TPR) und der Falsch-Positiv-Rate (engl. *False Positive Rate*; FPR) bei unterschiedlichen Schwellenwerten (Bradley, 1997; Zou et al., 2007). Die Fläche unter der Kurve (engl. *Area Under the Curve*; AUC) gibt an, wie gut das Modell insgesamt zwischen den beiden Klassen unterscheiden kann. Der ROC-AUC-Wert, also die Fläche unter der Grenzwertoptimierungskurve, beschreibt damit die Gesamtgüte der Klassifikation und berechnet sich als:

  $$ROC\ AUC = \int_0^1 TPR(FPR)\, dFPR$$

  mit $FPR = \frac{FP}{FP + TN}$

  Der Wertebereich liegt zwischen 0 und 1. Ein AUC-Wert von 1 bedeutet, dass das Modell eine perfekte Trennung zwischen den Klassen ermöglicht, während ein Wert von 0,5 auf eine zufällige Klassifikation hindeutet (Fawcett, 2006). Modelle mit einem höheren AUC-Wert haben eine bessere Fähigkeit, zwischen positiven und negativen Klassen zu unterscheiden.

1.2.5.9 Modellvergleichsanalyse

Zur Identifikation statistisch signifikanter Leistungsunterschiede zwischen den Klassifikationsmodellen wurde eine Modellvergleichsanalyse auf der Grundlage des Cochrane Q Tests durchgeführt (Cochran, 1950; Dietterich, 1998). Dieser auf der Chi Quadrat Verteilung beruhende Test erweitert den McNemar Test auf den Vergleich von mehr als zwei abhängigen dichotomen Klassifikationen, wie hier bei der Segmentzuordnung in die Klassen „iktal“ und „nicht-iktal“, und prüft, ob sich die Modelle in ihren Fehlklassifikationen systematisch unterscheiden (Sheskin, 2011). Die Nullhypothese geht davon aus, dass

alle Modelle eine vergleichbare Klassifikationsleistung aufweisen. Ein signifikanter Q Wert weist darauf hin, dass mindestens ein Modell von den übrigen Modellen abweicht.

Ergab der Cochrane Q Test einen signifikanten Befund, wurden ergänzend paarweise Post-hoc-Vergleiche mittels McNemar-Test durchgeführt (McNemar, 1947; Sheskin, 2011). Dieser nichtparametrische Test bewertet Unterschiede in den Fehlklassifikationen zweier Modelle anhand diskordanter Paare, also solcher Fälle, in denen die Modelle zu unterschiedlichen Klassifikationsurteilen gelangen. Die Auswertung erfolgt auf der Basis einer Vierfeldertafel und nutzt ebenfalls die Chi Quadrat Verteilung (Edwards, 1948). Ein signifikanter McNemar Test deutet darauf hin, dass ein Modell im Vergleich zu dem jeweils betrachteten Referenzmodell systematisch bessere oder schlechtere Vorhersagen liefert.

Die praktische Umsetzung der Modellvergleiche erfolgte mit der `MLxtend` Bibliothek (Raschka, 2018). Für alle Tests wurde ein Signifikanzniveau von 0,01 festgelegt. Eine detaillierte Darstellung der zugrunde liegenden Implementierung findet sich in Supplement B.4 der Originalpublikation (Kayser et al., 2025). Die in Abbildung 2 dargestellte methodische Kernimplementierung veranschaulicht den Ablauf der paarweisen McNemar Analysen.

1.2.5.10 Anfallserkennung

Nach Abschluss der Modellvergleichsanalyse wurde der leistungsstärkste Klassifikationsalgorithmus als Referenzmodell ausgewählt. Die Entscheidung basierte auf einer integrativen Bewertung aller zuvor erläuterten Metriken sowie den Ergebnissen des Cochrane-Q-Tests und der paarweisen McNemar-Analysen. Das final trainierte Modell wurde anschließend eingesetzt, um für jedes EEG-Segment eine probabilistische Klassenzuweisung zu berechnen. Die Hyperparameter des Klassifikators entsprachen den im Rahmen der Grid-Search-Optimierung bestimmten Konfigurationen (siehe Tabelle 2 und Abschnitt 1.2.5.7).

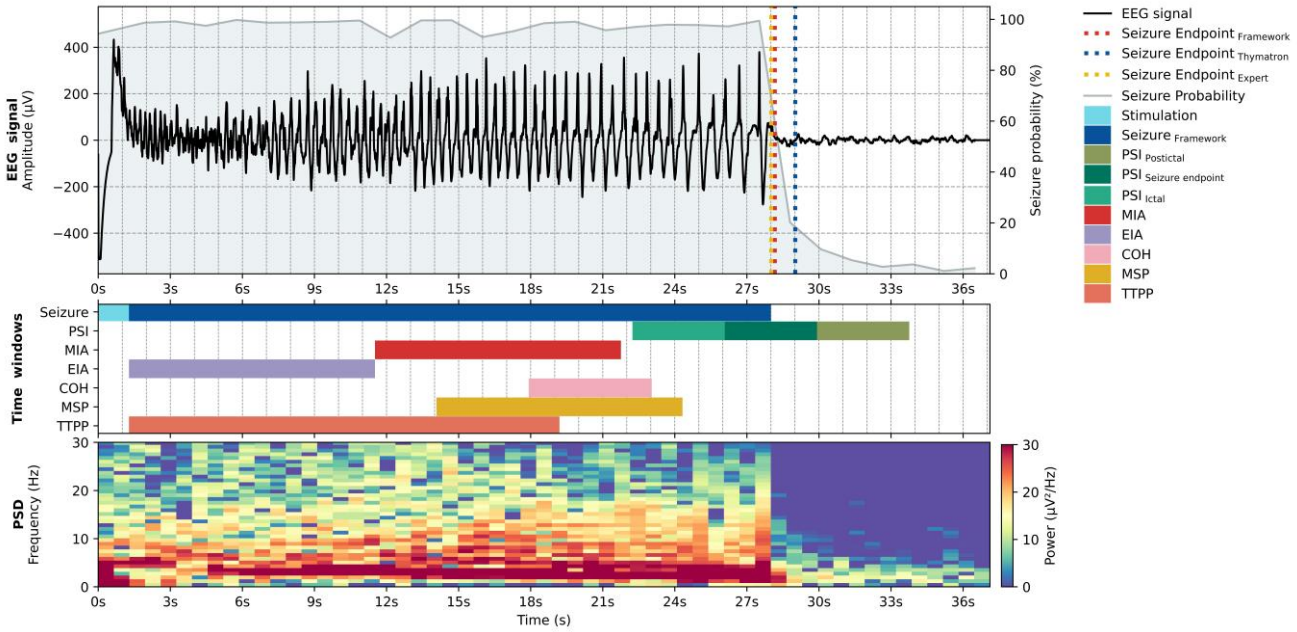

Abb. 2: Beispiel eines EKT-induzierten Krampfanfalls und seiner automatisierten Auswertung im entwickelten Analyse-Framework (modifiziert nach Kayser et al., 2025). Das obere Panel zeigt das Roh-EEG-Signal (linke y-Achse, Amplitude in µV) sowie die vom Klassifikationsmodell berechnete Anfallswahrscheinlichkeit (rechte y-Achse, grau). Die farbigen vertikalen Linien markieren das Krampfende gemäß Framework-Detektion (dunkelblau), dem Stimulationsgerät (Thymatron®) (orange) und der Expertenbewertung (rot). Das mittlere Panel visualisiert das Stimulationsereignis (hellblau) und das darauffolgende Krampfgeschehen (dunkelblau) sowie die zugrunde liegenden Zeitfenster der berechneten Qualitätsparameter (PSI, MIA, EIA, COH, MSP, TTPP). Das untere Panel zeigt die zeitaufgelöste Leistungsspektraldichte (engl. *Power Spectral Density*; PSD) des EEG-Signals. Höhere Leistungswerte sind durch wärmere Farben dargestellt.

In der Anwendung auf die vollständige EEG-Datenbasis wurde für jedes EEG-Segment die Wahrscheinlichkeit einer Zugehörigkeit zu den Klassen „iktal“, oder „nicht-iktal“ bestimmt. Der algorithmische Ablauf dieses Verfahrens ist in Abbildung 2 schematisch dargestellt. Der Anfallsbeginn wurde als der früheste Zeitpunkt nach der Stimulation definiert, zu dem die vorhergesagte Wahrscheinlichkeit für die Klasse „iktal“ für eine Dauer von mindestens 8 Sekunden die Schwelle von 50 % überschritt. Das Anfallsende wurde als jener Zeitpunkt festgelegt, zu dem die Wahrscheinlichkeit für „iktal“ in drei unmittelbar aufeinanderfolgenden Segmenten unter 30 % lag. In diesem Fall wurde das erste Segment, das als „nicht-iktal“ klassifiziert wurde, als Anfallsendpunkt gewertet. Trat innerhalb eines

Zeitfensters von vier Sekunden nach diesem Zeitpunkt erneut ein Anstieg der Wahrscheinlichkeit für die Klasse iktal auf und erfüllten die nachfolgenden Segmente erneut die genannten Kriterien für einen Anfall, wurde das spätere Ende als gültiger Anfallsendpunkt übernommen. Diese Regelung sollte verfrühte Anfallsbeendigungen vermeiden und erlaubte die Erfassung postiktaler Wiederholungsphänomene, wie sie unmittelbar im Anschluss an einen Anfall auftreten können.

1.2.6 Berechnung der Qualitätsparameter

Im Anschluss wurden aus den EEG-Signalen verschiedene Qualitätsparameter der Elektrokonvulsionstherapie berechnet. Diese Parameter erfassen zentrale Charakteristika der iktalen Hirnaktivität, der postiktalen EEG-Veränderungen und der neuronalen Synchronisation (Azuma et al., 2007; Krystal et al., 1995; Suppes et al., 1996). Sie basieren auf der zuvor bestimmten Anfallsdauer und wurden in Anlehnung an etablierte gerätebasierte Verfahren definiert (Somatics LLC, 2019). Die vom Stimulationsgerät automatisch berechneten Werte dienten als Referenz zur Validierung der eigenen Berechnungen.

Abbildung 2 gibt eine schematische Übersicht über die Segmentierung der EEG-Signale, die Zuordnung zu iktalen und postiktalen Phasen sowie die zur Berechnung der einzelnen Krampfqualitätsmetriken herangezogenen Zeitfenster und Frequenzbereiche. Zur Bestimmung spektraler Parameter wurden die EEG-Signale frequenzbasiert zerlegt. Das Leistungsdichtespektrum wurde in mehreren physiologisch und neurophysiologisch relevanten Frequenzbändern berechnet, die unterschiedliche funktionelle Zustände des Gehirns widerspiegeln. Ausgewertet wurde die Delta-Band-Aktivität im Bereich von 0,7 bis 3,5 Hz, die Theta-Band-Aktivität von 3,5 bis 8 Hz, die Alpha-Band-Aktivität von 8 bis 13 Hz und die Beta-Band-Aktivität von 13 bis 25 Hz. Die Berechnung erfolgte mithilfe einer schnellen Fourier-Transformation mit Funktionen aus der `SciPy` Bibliothek (Duhamel und Vetterli, 1990; Virtanen et al., 2020). Hierzu wurde auf nicht überlappende Segmente jeweils eine Hanning-Fensterfunktion angewendet.

Der zur Berechnung der Qualitätsparameter entwickelte Python-Code wurde in der Toolbox `ECTMetrics` zusammengeführt und als Open-Source-Projekt veröffentlicht (Kayser, 2025).

Im Folgenden werden die einzelnen Qualitätsparameter und ihre Berechnung näher erläutert.

1.2.6.1 Krampfdauer

Die Krampfdauer (engl. *seizure duration; SE*) wurde als Zeitintervall zwischen Anfallsbeginn und Anfallsende ausschließlich anhand des EEG-Signals bestimmt, wie es durch das in Abschnitt 1.2.5.10 beschriebene Verfahren definiert und in der nachfolgenden Gleichung formal dargestellt ist. Während die Krampfdauer traditionell als Marker für die Anfallsauslösung galt, wird ihre Aussagekraft für den Therapieerfolg der EKT inzwischen differenzierter bewertet. Aktuelle Studien zeigen zwar eine Assoziation zwischen Krampfdauer und Remission, machen jedoch deutlich, dass weitere klinische Faktoren den Behandlungserfolg wesentlich mitbestimmen (Gillving et al., 2024). Im Vergleich zur motorischen oder EMG basierten Erfassung gilt die EEG basierte Messung jedoch weiterhin als robusteres Verfahren, da sie das neuronale Anfallsgeschehen direkter abbildet (Sørensen et al., 1981). Als klinisch ausreichend gelten EEG-erfasste Krampfanfälle mit einer Dauer von mindestens 25Sekunden; während bei Patienten über 75Jahren bereits eine Dauer von 20Sekunden als adäquat bewertet werden kann (Grözinger et al., 2013).

$$\text{Krampfdauer}\ (s) = t_{Anfallsende} - t_{Anfallsbeginn}$$

mit $t_{\text{Anfallsbeginn}} < t_{\text{Anfallsende}}$

wobei $t_{\text{Anfallsbeginn}}$ bezeichnet den Zeitpunkt des Anfallsbeginns und $t_{\text{Anfallsende}}$ den Zeitpunkt des Anfallsendes.

Krampfanfälle mit einer Dauer von weniger als 10Sekunden wurden von der weiteren Analyse ausgeschlossen, da die zugrunde liegenden Berechnungsverfahren mehrerer Qualitätsparameter für derart kurze EEG-Abschnitte per Definition nicht anwendbar sind.

1.2.6.2 Postiktaler Suppressionindex

Der postiktale Suppressionsindex (engl. *Postictal Suppression Index*; PSI) quantifiziert den Übergang von der iktalen Aktivität in die abgeschwächte postiktale EEG-Phase und wird als Maß für die zentrale Inhibitionsleistung interpretiert.

Für die vorliegende Arbeit wurde der PSI als relative Reduktion der mittleren EEG-Amplitude im postiktalen im Vergleich zum iktalen Abschnitt definiert. Hierzu wurden je drei EEG-Segmente im postiktalen und im iktalen Zeitraum herangezogen. Um unmittelbare Übergangseffekte am Anfallsendpunkt zu vermeiden, wurde ein Zeitfenster von vier Sekunden, entsprechend drei Segmenten, um den Anfallsendpunkt von der Analyse ausgeschlossen (Baghai et al., 2004). Aus den verbleibenden Segmenten wurden die mittleren absoluten Amplituden für den iktalen und den postiktalen Abschnitt berechnet.

$$PSI\ (\%) = \left(1 - \frac{\sum_{i=1}^{3} \left|A_{(STS\,+\,1)\,+\,i}\right|}{\sum_{i=1}^{3} \left|A_{(STS\,-\,1)\,-\,i}\right|}\right) \times 100$$

mit $PSI \in [0; 100]$

wobei STS das Segment des Anfallsendes (engl. *seizure termination segment*) bezeichnet. $\left|A_{(STS\,+\,1)\,+\,i}\right|$ ist die absolute EEG-Signalamplitude des i –ten postiktalen Segments, während $\left|A_{(STS\,-\,1)\,-\,i}\right|$ die absolute EEG-Signalamplitude des i-ten präterminalen iktalen Segments angibt.

Der PSI wird als Prozentwert angegeben und liegt im Bereich von 0 bis 100 %. Studien zeigten Hinweise auf eine positive Korrelation zwischen dem PSI und dem Behandlungserfolg (Azuma et al., 2007; Nobler et al., 2000; Nobler et al., 1993; Suppes et al., 1996). PSI Werte über 80 % gelten daher als ideal (Grözinger et al., 2013). Bei niedrigeren Werten kann eine Restimulation erwogen werden (Baghai et al., 2004; Weiner et al., 1991). Tritt im vermeintlich postiktalen EEG-Abschnitt eine verstärkte Erregung auf, kann der PSI rechnerisch negative Werte annehmen. In der vorliegenden Arbeit wurden solche Verläufe als Ausdruck einer fehlenden Suppression interpretiert und der PSI konsequent auf 0 % gesetzt.

1.2.6.3 Konvulsions-Energie-Index

Der Konvulsions-Energie-Index (engl. *Average Seizure Energy Index*; ASEI) beschreibt die mittlere Leistung des EEG-Signals während eines Krampfanfalls und dient als Maß für die energetische Intensität der neuronalen Aktivität (Baghai et al., 2004). Er wurde berechnet, indem die quadrierte EEG-Amplitude über die gesamte Krampfdauer gemittelt wurde.

$$\text{ASEI } (\mu V^2) = \frac{1}{Krampfdauer} \int_{t_{Anfallsbeginn}}^{t_{Anfallsende}} P(t)dt$$

mit $P(t) = x(t)^2$ und $t_{Anfallsbeginn} < t_{Anfallsende}$

wobei $x(t)$ die EEG-Amplitude zum Zeitpunkt t, während $P(t)$ das Leistungsdichtespektrum des EEG-Signals darstellt. Die Krampfdauer entspricht dem Zeitintervall zwischen Anfallsbeginn $t_{Anfallsbeginn}$ und Anfallsende $t_{Anfallsende}$ wie in Abschnitt 1.2.6.1 beschrieben.

1.2.6.4 Maximal anhaltende Leistung

Die maximal anhaltende Leistung (engl. *Maximum Sustained Power*; MSP) beschreibt die höchste kumulierte EEG-Leistung innerhalb eines Zeitraums von acht aufeinanderfolgenden iktalen EEG Segmenten, was einem Zeitfenster von etwa 10 Sekunden während des Anfalls entspricht (Rasmussen, 2002; Somatics LLC, 2019). Für jedes mögliche Fenster aus acht konsekutiven iktalen Segmenten wurde die Summe der EEG-Leistung berechnet, und der höchste dieser Werte wurde als MSP definiert. Hohe MSP Werte geben Aufschluss über die maximal anhaltende neuronale Erregung während eines Krampfanfalls und könnten mit der therapeutischen Wirksamkeit der Elektrokonvulsionstherapie in Zusammenhang stehen (Francis-Taylor et al., 2020).

$$MSP\left(\frac{\mu V^2}{Hz}\right) = \max_{\mathrm{SOS} \le \mathrm{k} < (\mathrm{STS}-8)} \left(\frac{1}{8}\sum_{i=0}^{7} P_{k+i}\right)$$

mit $SOS < (STS - 7)$

wobei SOS bezeichnet das Segment zum Zeitpunkt des Anfallsbeginns, STS das Segment zum Zeitpunkt des Anfallsendes. Der Index k gibt den Startpunkt eines Fensters aus acht aufeinanderfolgenden iktalen Segmenten an, der Index i die Position innerhalb dieses Fensters. P_{k+i} entspricht der mittleren EEG-Leistung im $(k + i)$-ten Segment.

1.2.6.5 Zeit bis zur maximalen Leistung

Die Zeit bis zur maximalen Leistung (engl. *Time To Peak Power*; TTPP) bezeichnet das Zeitintervall zwischen dem Ende des Stimulus und dem Zeitpunkt, an dem die maximale EEG-Leistung während eines Krampfanfalls erreicht wird (Somatics LLC, 2019). Dieser Wert gibt Aufschluss darüber, wie schnell eine maximale neuronale Aktivierung erfolgt. Eine kurze TTPP weist auf eine rasche Rekrutierung kortikaler Strukturen hin, während längere Werte eine verzögerte maximale Krampfaktivität widerspiegeln können (Sackeim, 1996).

$$TTPP\ (s) = t_{MSP} - t_{Anfallsbeginn}$$

mit $t_{MSP} \geq (t_{Anfallsbeginn} + 5.12)$

wobei t_{MSP} bezeichnet den Zeitpunkt der maximal anhaltenden Leistung, und $t_{Anfallsbeginn}$ den Zeitpunkt des Anfallsbeginns.

1.2.6.6 Frühe iktale Amplitude

Die frühe iktale Amplitude (engl. *Early-Ictal Amplitude*; EIA) bildet die initiale Krampfintensität unmittelbar nach der Stimulation ab. Höhere frühe- und mittel-iktale Amplituden

(siehe auch Abschnitt 1.2.6.7) sind mit Merkmalen adäquater Reizintensität und generalisierter Anfallsaktivität assoziiert und fließen auch in Modelle zur Beurteilung der Adäquanz der Stimulation ein (De Arriba-Arnau et al., 2021; Francis-Taylor et al., 2020; Krystal et al., 2000; Krystal et al., 1995; Krystal et al., 1993). Die EIA wurde als mittlere absolute EEG-Amplitude der ersten acht iktalen Segmente berechnet.

$$\text{EIA } (\mu V) = \frac{1}{8} \sum_{i=SOS}^{SOS+7} |A_i|$$

mit $SOS < (STS - 7)$

wobei SOS bezeichnet das Segment zum Zeitpunkt des Anfallsbeginns, STS das Segment zum Zeitpunkt des Anfallsendes, und $|A_i|$ die absolute EEG-Amplitude im i-ten initialen iktalen Segment.

1.2.6.7 Mittlere iktale Amplitude

Die mittlere iktale Amplitude (engl. *Mid-Ictal Amplitude*; MIA) beschreibt die höchste mittlere Amplitude innerhalb eines Zeitraums von acht aufeinanderfolgenden Segmenten während der Anfallsperiode. Ein hoher MIA-Wert weist auf eine intensive neuronale Entladung hin und ist mit Merkmalen einer adäquaten Generalisierung des Krampfgeschehens assoziiert (Janouschek et al., 2020; Krystal et al., 1998). Niedrige MIA-Werte können hingegen auf eine unzureichende Synchronisation kortikaler Aktivität oder eine vorzeitige Anfallsbeendigung hinweisen (Krystal et al., 2000). Eine mittlere iktale Amplitude von mindestens 180 µV, beziehungsweise 150 µV bei Patienten über 75 Jahren, wird als ausreichend angesehen (Grözinger et al., 2013). Die MIA wurde als maximaler Mittelwert der absoluten EEG-Amplituden über gleitende Fenster von acht konsekutiven iktalen Segmenten berechnet.

$$MIA\ (\mu V) = max_{SOS \leq k < (STS-8)} \left(\frac{1}{8} \sum_{i=0}^{7} |A_{k+i}| \right)$$

mit $SOS < (STS - 7)$

wobei SOS bezeichnet das Segment zum Zeitpunkt des Anfallsbeginns, STS das Segment zum Zeitpunkt des Anfallsendes, k den Startindex eines Fensters aus acht aufeinanderfolgenden iktalen Segmenten, i den Index innerhalb dieses Fensters und $|A_{k+i}|$ die absolute EEG-Amplitude im $(k+i)$-ten iktalen Segment.

1.2.6.8 Maximale interhemisphärische Kohärenz

Die maximale interhemisphärische Kohärenz (engl. *Maximum Sustained Coherence*; COH), auch Konvulsions-Konkordanz-Index genannt, erfasst die interhemisphärische Synchronisation des EEG während eines Krampfanfalls. Als paarweises Maß erfordert die COH je eine links- und rechtshemisphärische EEG-Ableitung (Shaw, 1981). Studien zeigen Zusammenhänge zwischen höheren iktalen Kohärenzwerten und Merkmalen adäquater Anfallsaktivität bzw. klinischer Effekte (Krystal et al., 1995; Perera et al., 2004). Längsschnittliche Analysen zeigen steigende iktale Kohärenzwerte im Behandlungsverlauf, ohne dass diese Zunahme die antidepressive Wirksamkeit allein erklären würde (Ten Doesschate et al., 2019). Darüber hinaus deuten Vorbefunde darauf hin, dass präther-apeutische interhemisphärische Kohärenz mit dem späteren EKT Ansprechen assoziiert sein kann (Roemer et al., 1990). In Anlehnung an Grözinger und Kollegen wird eine Synchronizität von mindestens 90 % als suffizient gewertet (Grözinger et al., 2013).

Die COH wurde als höchste mittlere Kohärenz innerhalb eines etwa 10 Sekunden langen iktalen Intervalls bestimmt. Hierzu wurden gleitende Fenster aus acht konsekutiven iktalen Segmenten betrachtet. Für jedes dieser Fenster wurde der Mittelwert der Kohärenzwerte der einzelnen Segmente berechnet und der höchste dieser Werte als maximale interhemisphärische Kohärenz definiert.

$$COH\ (\%) = max_{SOS \leq k < (STS-8)} \left(\frac{1}{8} \sum_{i=0}^{7} C_{k+i} \right)$$

mit $SOS < (STS - 7)$

wobei SOS bezeichnet das Segment zum Zeitpunkt des Anfallsbeginns, STS das Segment zum Zeitpunkt des Anfallsendes, k

den Startindex eines Fensters aus acht aufeinanderfolgenden iktalen Segmenten, und C_{k+i} die Kohärenz im $(k+i)$-ten iktalen Segment. Die Kohärenz C wurde aus der quadrierten Kreuzspektraldichte $|P_{links-rechts}|^2$ der beiden hemisphärischen Signale und den entsprechenden Leistungsdichtespektren P_{links} und P_{rechts} der linken und rechten Ableitung berechnet nach: $C = \frac{|P_{links-rechts}|^2}{P_{links} \times P_{rechts}}$

1.2.6.9 Statistische Analyse

Die statistische Analyse gliederte sich in eine Voraussetzungsprüfung auf Verteilungsform und Homoskedastizität sowie anschließende inferenzstatistische und reliabilitätsbezogene Auswertungen.

Die Normalverteilung der Daten wurde mithilfe des Shapiro-Wilk-Tests überprüft (Shapiro und Wilk, 1965). Fiel der Test signifikant aus und zeigte damit eine Abweichung von der Normalverteilung, wurden nichtparametrische Verfahren verwendet. Ergänzend wurde die Homoskedastizität mit dem Levene-Test geprüft, um die Gleichheit der Gruppenvarianzen abzusichern (Levene, 1960). Da viele parametrische Verfahren Varianzhomogenität voraussetzen, diente das Ergebnis als Grundlage für die Wahl zwischen parametrischen und varianzunabhängigen nichtparametrischen Methoden. Für beide Tests galt $\alpha < 0{,}01$ als Signifikanzniveau.

Zur Bestimmung des Zusammenhangs zwischen berechneten Anfallsqualitätsindizes, Expertenbewertungen (Anfallsdauer und PSI) und gerätegenerierten Werten wurde der Spearman-Rangkorrelationskoeffizient berechnet (Spearman, 1904). Dieser Korrelationskoeffizient eignet sich insbesondere für nicht-normalverteilte oder ordinal skalierte Daten, da er anstelle linearer Zusammenhänge die monotone Beziehung zwischen zwei Variablen beschreibt. Ein Spearman-Koeffizient nahe +1 zeigt eine starke positive Korrelation zwischen den Variablen. Wert nahe -1 sprechen für einen inversen Zusammenhang, während Werte nahe 0 auf eine fehlende Rangkorrelation hindeuten. Zur Untersuchung von Gruppenunterschieden wurde zusätzlich der Mann-Whitney-U-Test (MWU), auch bekannt als

Wilcoxon-Rangsummentest, angewandt (Mann und Whitney, 1947). Es handelt sich dabei um ein nicht-parametrisches Verfahren, das Unterschiede zwischen zwei unabhängigen Gruppen auf Basis der Rangverteilung testet. Da der Test keine Normalverteilung voraussetzt, eignet er sich besonders für Vergleiche zwischen den berechneten Parametern und den Referenzwerten des Stimulationsgeräts sowie für Subgruppenanalysen innerhalb der Stichprobe. Für beide Tests wurde eine statistische Signifikanz von $\alpha < 0{,}01$ als Schwellenwert definiert.

Um die Übereinstimmung zwischen den berechneten Anfallsqualitätsindizes, den Expertenbewertungen und den gerätebasierten Messungen zu quantifizieren, wurde der Intraklassen-Korrelationskoeffizient (engl. *Intraclass Correlation Coefficient*; ICC) anhand der `Pingouin` Toolbox (Vallat, 2018) bestimmt. Der ICC bewertet die Konsistenz von Messwerten, indem er die Varianz zwischen und innerhalb den Messmethoden ins Verhältnis setzt. Werte nahe +1 weisen auf eine exzellente Übereinstimmung hin, während Werte unter 0,5 auf eine geringe Reliabilität hindeuten. Fehlende Werte wurden mittels paarweisen Ausschlusses behandelt, um die Vergleichbarkeit der Analysen zu maximieren. Die Schätzung erfolgte anhand eines zweifaktoriellen Mischmodells, welches sowohl systematische Unterschiede zwischen den Beurteilern als auch zufällige Messfehler berücksichtigt. Zur Bewertung der individuellen Übereinstimmung einzelner Messverfahren wurde der ICC_3 berechnet, während der ICC_3K zur Abschätzung der durchschnittlichen Übereinstimmung über mehrere Messmethoden hinweg eingesetzt wurde (Liljequist et al., 2019; McGraw und Wong, 1996; Shrout und Fleiss, 1979). Zur statistischen Absicherung wurde die Signifikanz mittels eines F-Tests überprüft, wobei ein Signifikanzniveau von $\alpha < 0{,}01$ festgelegt wurde.

1.3 Ergebnisse

1.3.1 Datensatz und klinische Charakteristika

Einen Überblick über die demographischen und allgemeinen klinischen Daten der Studie gibt Tabelle 1. Der analysierte Datensatz umfasste insgesamt 1634 Elektrokonvulsionstherapie EKT-Behandlungen, die an 116 Patienten durchgeführt wurden. Von diesen waren 75 weiblich (64,7 %) und 41 männlich (35,3 %). Das Durchschnittsalter der Patienten

lag bei 59 Jahren (Spannweite: 19 bis 90 Jahre). Die Anzahl der durchgeführten Behandlungen pro Patienten variierte, wobei mindestens 5 und maximal 29 Sitzungen dokumentiert wurden. Die mittlere Anzahl an EKT-Sitzungen pro Patienten betrug 14 ± 6. Von den insgesamt 116 Patienten erfüllten 83 Patienten (71,6 %) die vordefinierten Kriterien für ein Therapieansprechen (Responder), während 33 Patienten (28,4 %) als Nicht-Responder eingestuft wurden. In Bezug auf die diagnostische Verteilung wies die Mehrheit der Patienten eine rezidivierende depressive Störung, aktuelle Episode schwer ohne psychotische Symptome (ICD-10: F33.2; n = 67; 57,8 %), auf. Weitere häufige Diagnosen waren eine rezidivierende depressive Störung mit psychotischen Symptomen (F33.3; n = 14; 12,1 %) sowie eine mittelgradige depressive Episode (F32.2; n = 12; 10,3 %). Eine schwere depressive Episode mit psychotischen Symptomen (F32.3) lag bei 8 Patienten (6,9 %) vor; bei weiteren 8 Patienten (6,9 %) wurde eine bipolare affektive Störung mit gegenwärtiger depressiver Episode (F31.4) diagnostiziert. Seltener traten andere affektive Störungen auf (F31.8, F33.1, F33.4; jeweils ≤ 5 Patienten). Eine unilateral rechte Stimulation wurde mit 70 Fällen (60,3 %) am häufigsten verwendet, gefolgt von einer bifrontalen Platzierung, die bei 30 Patienten (25,9 %) zum Einsatz kam. Eine bitemporale Elektrodenanordnung wurde bei 10 Patienten (8,6 %) angewendet, während eine unilateral linke Stimulation nur bei 5 Patienten (4,3 %) erfolgte. Eine links-anterior-rechts-temporale Elektrodenanordnung wurde lediglich bei einem Patienten (0,9 %) dokumentiert. Die während der Behandlung verabreichte elektrische Ladung zeigte eine große Spannweite zwischen 25 mC und 1008 mC mit einem Mittelwert von 370 ± 215 mC. Bei Patienten, die ein Therapieansprechen zeigten, lag die mittlere Stimulationsladung bei 360 ± 210 mC, während sie bei den Nicht-Respondern mit 396 ± 226 mC etwas höher war. Die Stimulationsenergie betrug im Durchschnitt 73 ± 42 % der maximalen Geräteeinstellung, mit einer dokumentierten Bandbreite von 5 % bis 200 %. Die Stimulationsdauer variierte zwischen 3 und 8 Sekunden, mit einem Mittelwert von 7 ± 1 Sekunden. Die analysierten EEG-Aufzeichnungen zeigten eine variable Länge, abhängig von der individuellen Anfallsdauer. Die kürzeste Aufzeichnung betrug 5,12 Sekunden (4 Segmente), die längste 115,2 Sekunden (90 Segmente). Die durchschnittliche Dauer lag bei 40,52 Sekunden (31,7 Segmente), der Median bei 37,12 Sekunden (29 Segmente). Acht EEG-Aufzeichnungen mit einer Gesamtdauer von über 120 Sekunden wurden ausgeschlossen (siehe Abschnitt 1.2.4).

Tab. 1: Klinische Charakteristika und Behandlungsparameter der Studienkohorte (modifiziert nach Kayser et al., 2025).

				Summe	Responder	Non-Responder	Minimum	Maximum	Einheit
Klinische Charakteristika	Patienten			116	83	33			n
	Geschlecht		Männlich	75	57	18			n
			Weiblich	41	26	15			n
	Alter			59 (16)	61 (16)	55 (16)	19	90	Mittelwert (SD)
	Gewicht			75 (20)	73 (20)	80 (18)	44	176	Mittelwert (SD), kg
	Diagnose	Bipolare Störung, derzeit depressive Episode	F31.4	8	8	0			n
			F31.8	1	1	0			n
		Schwere depressive Störung	F32.2	12	7	5			n
			F32.3	8	8	0			n
			F33.1	5	4	1			n
			F33.2	67	42	25			n
			F33.3	14	13	1			n
			F33.4	1	1	0			n
Behandlungsparameter	Anzahl an EKT-Sitzungen			14 (6)	15 (5)	13 (7)	5	29	Mittelwert (SD), n
	Stimulationsenergie			73 (42)	71 (41)	78 (44)	5	200	Mittelwert (SD), %
	Stimulus-Dauer			7 (1)	7 (1)	7 (1)	3	8	Mittelwert (SD), s
	Abgegebene Ladung			370 (215)	360 (210)	396 (226)	25	1022	Mittelwert (SD), mC
	Elektroden-Position		Bifrontal	30	25	5			n
			Bitemporal	10	5	5			n
			Links anterior rechts temporal	5	4	1			n
			Unilateral links	1	1	0			n
			Unilateral rechts	70	48	22			n

1.3.2 Segmentbasierte Anfallserkennung

Bei der Segmentierung der EEG-Daten wurden insgesamt 51.722 Segmente erzeugt. Davon wurden 12.198 Segmente (23,6 %) wegen fehlender Labels und 1.157 Segmente (2,2 %) aufgrund zeitlicher Überlappung mit der Stimulationsphase ausgeschlossen. Die verbleibenden 38.367 Segmente entfielen auf 28.622 (74,6 %) iktale und 9.745 (25,4 %) nicht-iktale Segmente. Nach der Unterabtastung, bei der 18.877 Segmente aus der Mehrheitsklasse entfernt wurden, umfassten beide Klassen („iktal“ und „nicht-iktal“) jeweils

9.745 Instanzen. Das resultierende Datenset mit insgesamt 19.490 EEG-Segmenten wurde anschließend in Trainings- (n_{train}= 15.592) und Testsatz (n_{test}= 3.898) aufgeteilt.

1.3.2.1 Merkmalsanalyse

Die analysierten Merkmale wiesen eine ausgeprägte Variabilität auf (siehe Tabelle S3 im Supplement; Kayser et al., 2025). Die absolute Energie reichte von 1,00 µV² bis 4,70 × 10^7 µV² (Mittelwert: 1,18 × 10^6 µV², Standardabweichung: 2,12 × 10^6 µV²), während die Fläche unter der Grenzwertoptimierungs-Kurve (ROC AUC) Werte zwischen 0,08 und 516,00 µV·s aufwies (Mittelwert: 51,70 µV·s, Standardabweichung: 49,20 µV·s). Die frequenzbezogenen Merkmale zeigten ebenfalls eine breite Streuung. Die fundamentale Frequenz variierte zwischen 0,00 und 50,40 Hz (Mittelwert: 2,12 Hz, Standardabweichung: 1,15 Hz), während die Medianfrequenz Werte von 0,00 bis 49,60 Hz erreichte (Mittelwert: 8,01 Hz, Standardabweichung: 3,80 Hz). Die maximale Frequenz reichte von 0,00 bis 95,30 Hz (Mittelwert: 51,00 Hz, Standardabweichung: 15,80 Hz). Die statistischen Signalmerkmale umfassten unter anderem die Kurtosis, die zwischen -3,00 und 251,00 schwankte (Mittelwert: 0,12, Standardabweichung: 2,50). Die mittlere absolute Differenz lag zwischen 0,00 und 127,00 µV (Mittelwert: 7,85 µV, Standardabweichung: 8,56 µV), während der Mittelwert der Signalamplitude Werte zwischen -140,00 µV und 144,00 µV erreichte (Mittelwert: 0,42 µV, Standardabweichung: 6,78 µV). Die spektralen Merkmale umfassten die Leistungsbandbreite, die zwischen 0,00 und 66,40 Hz lag (Mittelwert: 13,70 Hz, Standardabweichung: 7,65 Hz). Die Null-Durchgangsrate reichte von 0,00 bis 70,00 Hz (Mittelwert: 15,70 Hz, Standardabweichung: 8,69 Hz). Zusätzliche Merkmale zur Beschreibung der Signalstruktur waren die negativen Umkehrpunkte, die Werte zwischen 0 und 52 aufwiesen (Mittelwert: 20,00, Standardabweichung: 6,20), sowie die Nachbarschaftspeaks, die zwischen 0 und 18 variierten (Mittelwert: 6,42, Standardabweichung: 2,16).

Von den ursprünglich 34 extrahierten Merkmalen wurden 17 für die weitere Analyse beibehalten. Dazu zählten die absolute Energie, die Fläche unter der Kurve, die Fundamentalfrequenz, die Wölbung, die maximale Amplitude, die maximale Frequenz, der Mittel-

wert, die mittlere absolute Differenz zwischen benachbarten Werten, der Median, der medianer Wert der Differenzen, die mediane Frequenz, das Minimum, die negativen Umkehrpunkte, die benachbarten Signalspitzen, die Leistungsbandbreite, die Schiefe sowie die Nulldurchgangsrate. Abbildung 3 zeigt eine farbkodierte Korrelationsmatrix dieser beibehaltenen Merkmale. Mithilfe einer Hauptkomponentenanalyse wurde die Anzahl der Hauptkomponenten im Anschluss auf sechs reduziert.

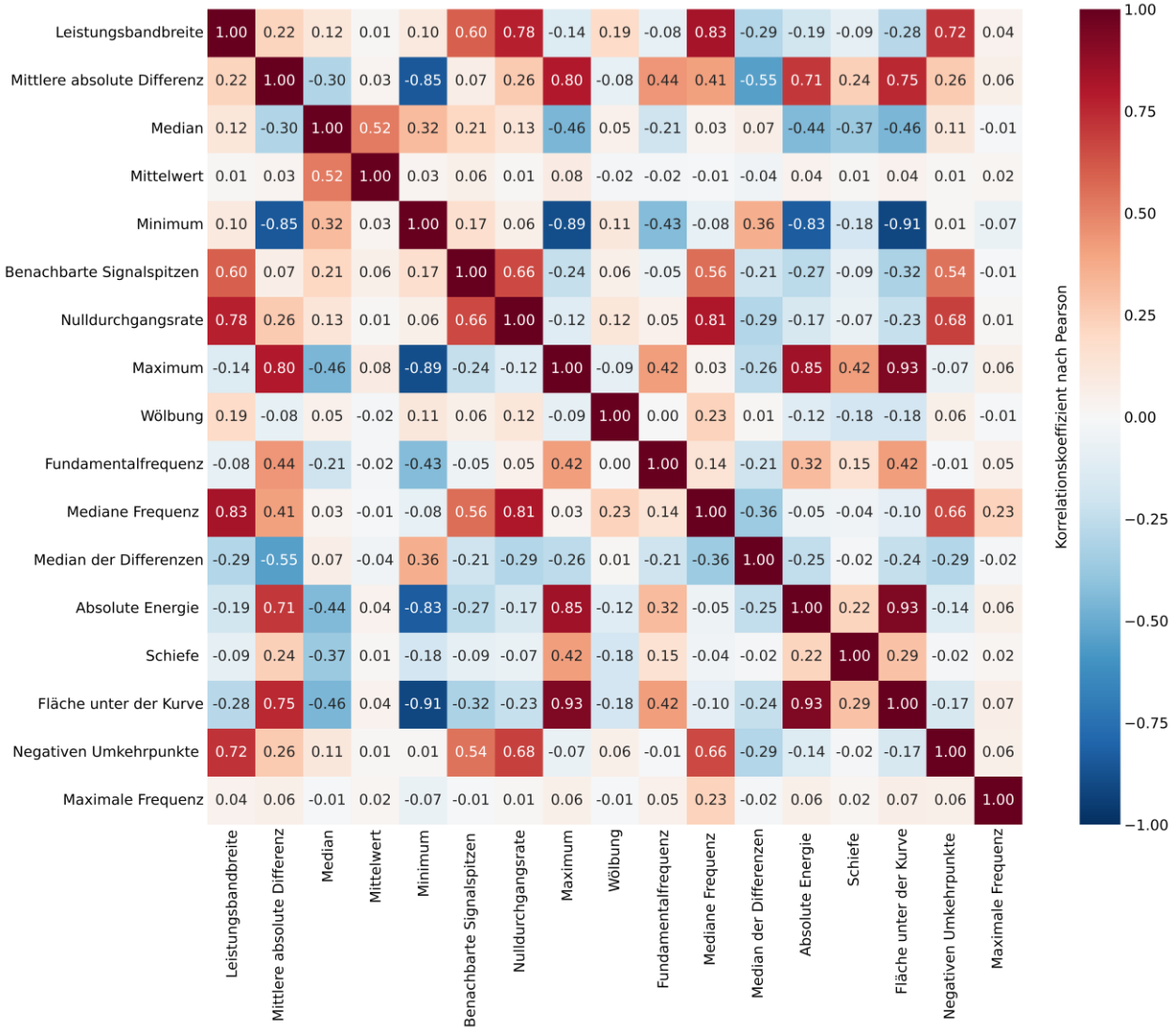

Abb. 3: Farbkodierte Korrelationsmatrix der extrahierten Merkmale nach der Merkmalsselektion. Dargestellt sind die paarweisen Korrelationskoeffizienten nach Pearson zwischen den für die segmentbasierte Anfallserkennung ausgewählten Merkmalen. Die Farbskala kodiert Richtung und Stärke der Korrelation.

1.3.2.2 Hyperparameteroptimierung

Insgesamt wurden über alle Klassifikatoren hinweg 6.705 Iterationen durchgeführt. Daraus ergaben sich für jeden Klassifikations-Algorithmus spezifische Hyperparameter Konfigurationen, mit der die jeweils höchste Klassifikationsleistung erreicht wurde (siehe Tabelle 2):

- Bei den Entscheidungsbäumen erwiesen sich eine maximale Tiefe von 10, das Entropie-Kriterium, eine Quadratwurzel-basierte Auswahl der Merkmale sowie jeweils acht erforderliche Beobachtungen pro Blatt und pro Split in Verbindung mit einer ausgeglichenen Klassengewichtung als optimal.
- Für Random Forests zeigten sich 75 Bäume, eine maximale Tiefe von 15, eine $\log_2$-transformierte Merkmalsauswahl, drei Beobachtungen pro Blatt und vier pro Split als beste Einstellung, ebenfalls mit ausgeglichener Klassengewichtung.
- Support Vector Machines erzielten die höchste Güte mit einem Kernel mit radialer Basisfunktion, einem Regularisierungsparameter C = 10, einem skalierten Gamma und einer ausgeglichener Klassengewichtung.
- Bei den k-Nearest-Neighbors-Modellen erwies sich eine Konfiguration mit sieben Nachbarn, Ball-Tree-Algorithmus, euklidischer Distanzmetrik und uniformer Gewichtung als optimal.
- Gradient Boosting zeigte die beste Leistung mit 125 Bäumen, einer maximalen Tiefe von fünf, einer Lernrate von 0,1 sowie der exponentiellen Verlustfunktion.

Tab. 2: Getestete Hyperparameter-Suchräume und optimale Parameterkombinationen der verwendeten Klassifikationsalgorithmen.

Algorithmus	Hyperparameter	Getestete Werte	Optimaler Wert
DT	Aufteilungskriterium	Gini-Index; Entropie	Entropie
	Maximale Tiefe	Unbegrenzt; 5; 10; 15	10
	Maximale Anzahl an Eingabemerkmalen	Quadratwurzel der Merkmalanzahl; Logarithmus zur Basis 2 der Merkmalanzahl	Quadratwurzel der Merkmalanzahl
	Minimale Stichprobengröße pro Blattknoten	5; 6; 7; 8	8
	Minimale Stichprobengröße für einen Split	6; 7; 8; 9	8
	Klassengewichtung	Ausgeglichen	Ausgeglichen
RF	Anzahl Bäume	75; 100; 125; 150	75
	Aufteilungskriterium	Gini-Index; Entropie	Entropie
	Maximale Tiefe	Unbegrenzt; 5; 10; 15	15
	Maximale Anzahl an Eingabemerkmalen	Quadratwurzel der Merkmalanzahl; Logarithmus zur Basis 2 der Merkmalanzahl	Logarithmus zur Basis 2 der Merkmalanzahl
	Minimale Stichprobengröße pro Blattknoten	1; 2; 3	3
	Minimale Stichprobengröße für einen Split	2; 4; 6; 8	4
	Klassen-Gewichtung	Ausgeglichen	Ausgeglichen
SVC	Regularisierungsparameter C	0,1; 1; 10	10
	Kernel-Funktion	Linear; Polynomial; Radial Basis Funktion	Radial Basis Funktion
	Gamma-Parameter	Kehrwert der Merkmalanzahl; Kehrwert von Merkmalanzahl × Gesamtvarianz des Trainings-Datensatzes; 0,001; 0,003; 0,005	Kehrwert von Merkmal-Anzahl × Gesamtvarianz des Trainingsdatensatzes
	Klassen-Gewichtung	Ausgeglichen	Ausgeglichen
KNN	Anzahl der Nachbarn	4; 5; 6; 7; 8; 9	7
	Algorithmus	K-Dimensionaler Baum; Ball-Tree	Ball-Tree
	Distanzmaß	Euklidisch; Manhattan; Chebyshev	Euklidisch
	Stimmgewichtung	Einheitlich; Distanzbasiert	Einheitlich
GBC	Anzahl der Boosting-Stufen	50; 75; 100; 125; 150	125
	Maximale Tiefe der Bäume	3; 4; 5; 6; 7	5
	Lernrate	0,01; 0,1; 0,2; 0,3	0,1
	Verlustfunktion	Exponentiell; Logistisch	Exponentiell

1.3.2.3 Modellvergleichsanalyse

Die Klassifikationsleistung der fünf hyperparametrisch optimierten Klassifikatoren ist in Tabelle 3 sowie Abbildung 4 und 5 dargestellt. RF, SVC und GBC erreichten jeweils 89 % Genauigkeit, Präzision, Recall und F1-Score, wobei RF mit einer ROC-AUC von 96 % und einem MCC von 0,80 die höchsten Werte aufwies. KNN erzielte in allen Metriken 88 % mit einem MCC von 0,77, während DT mit einer Genauigkeit von 86 %, einer ROC-AUC von 86 % und einem MCC von 0,72 die niedrigste Klassifikationsleistung zeigte.

Tab. 3: Klassifikationsgüte der trainierten Modelle für die segmentbasierte Anfallserkennung im unabhängigen Testdatensatz (modifiziert nach Kayser et al., 2025). Übersicht der Leistungsmetriken der fünf hyperparametrisch optimierten Klassifikatoren DT, RF, SVC, KNN und GBC. Dargestellt sind die Werte der Konfusionsmatrix (TP, TN, FP, FN) sowie die daraus berechneten Klassifikationsmetriken Accuracy, MCC, ROC AUC, F1-Score, Precision und Sensitivität (Recall). Die Werte beziehen sich auf die jeweilige Testmenge der segmentbasierten binären Anfallsklassifikation („iktal“ vs. „nicht-iktal“).

			Algorithmus				
			DT	**RF**	**SVC**	**KNN**	**GBC**
Konfusionsmatrix		**TP**	1672	1713	1700	1708	1712
		TN	1694	1767	1795	1755	1775
		FP	299	226	198	238	218
		FN	233	192	205	197	193
Metrik	**Accuracy**		0,86	0,89	0,90	0,89	0,89
	MCC		0,73	0,79	0,79	0,78	0,79
	ROC AUC		0,86	0,89	0,90	0,89	0,89
	F1-Score	*iktal*	0,86	0,89	0,90	0,89	0,90
		nicht-iktal	0,86	0,89	0,89	0,89	0,89
		Mittelwert	0,86	0,89	0,90	0,89	0,89
	Precision	*iktal*	0,88	0,90	0,90	0,90	0,90
		nicht-iktal	0,85	0,88	0,90	0,88	0,89
		Mittelwert	0,86	0,89	0,90	0,89	0,89
	Sensitivität	*iktal*	0,85	0,89	0,90	0,88	0,89
		nicht-iktal	0,88	0,90	0,89	0,90	0,90
		Mittelwert	0,86	0,89	0,90	0,89	0,89
Stützmenge		*iktal*	1993	1993	1993	1993	1993
		nicht-iktal	1905	1905	1905	1905	1905

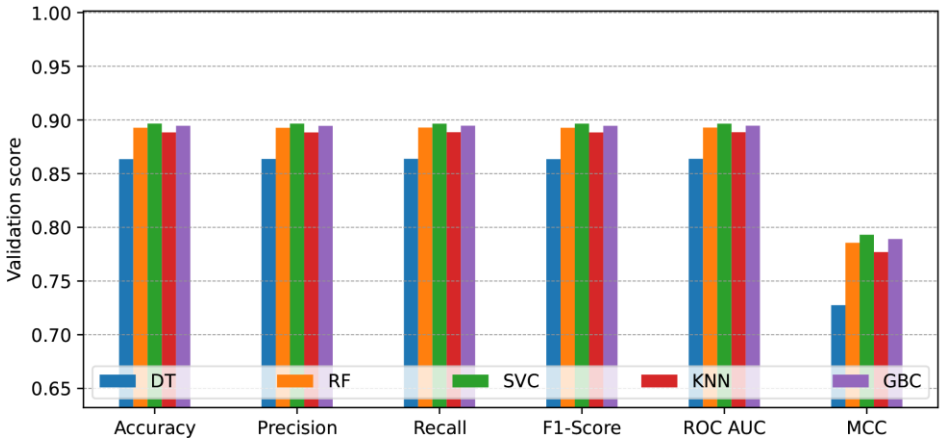

Abb. 4: Grafische Darstellung der Klassifikationsgüte der trainierten Modelle im unabhängigen Testdatensatz (modifiziert nach Kayser et al., 2025). Dargestellt sind die Metriken Accuracy, MCC, ROC AUC, F1, Precision und Sensitivität für die Modelle DT, RF, SVC, KNN und GBC.

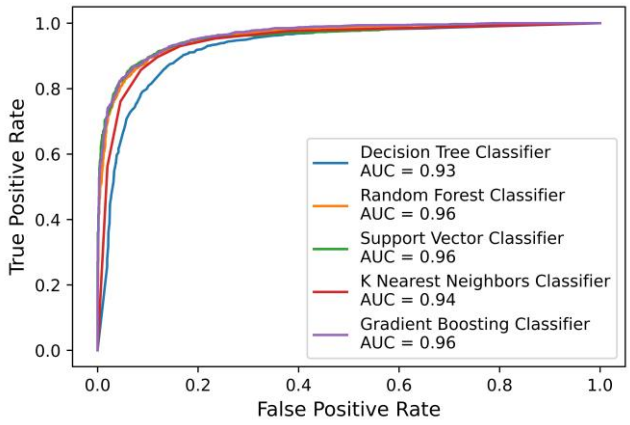

Abb. 5: ROC Kurven der trainierten Modelle im unabhängigen Testdatensatz (modifiziert nach Kayser et al., 2025). Dargestellt sind die Verläufe von TPR über FPR sowie die zugehörigen AUC-Werte für die Modelle DT, RF, SVC, KNN und GBC.

Die Unterschiede zwischen den Modellen in der Klassifikation von „iktalen“ und „nicht-iktalen“ Segmenten waren gering. Die höchsten Abweichungen in Präzision, Recall und

F1-Score zwischen den beiden Klassen betrugen 1 % für RF, SVC und GBC, 2 % für KNN und 3 % für DT. Die statistische Analyse der Gesamtklassifikationsleistung mittels Cochran-Q-Test ergab Q = 0,33, p = 0,85, womit keine signifikanten Unterschiede zwischen den Modellen nachgewiesen werden konnten. In den paarweisen Vergleichen mittels McNemar-Tests zeigte sich jedoch eine signifikant schlechtere Leistung von DT gegenüber allen anderen Modellen (χ^2 = 26,41 bis 46,68, alle p ≤ 0,01). Zudem war KNN signifikant schlechter als SVC (χ^2 = 7,72, p ≤ 0,05), während alle anderen Modellvergleiche keine signifikanten Unterschiede aufwiesen (p > 0,05). Die Ergebnisse der Modellvergleichsanalyse sind in Tabelle 4 zusammengefasst. Basierend auf diesen Ergebnissen wurde RF als Referenzmodell für die finale Klassifikationsaufgabe ausgewählt.

Tab. 4: Ergebnisse der paarweisen McNemar Tests zum Vergleich der Klassifikationsleistung der fünf trainierten Modelle (modifiziert nach Kayser et al., 2025). Dargestellt sind x^2 und p Werte der McNemar Tests für alle Modellpaare, wobei die Modellzeile das Referenzmodell und die Modellspalte das Vergleichsmodell bezeichnet. Diagonale Einträge bleiben leer, p Werte unter 0,01 gelten als statistisch signifikant.

		Vergleichsmodell				
		DT	**RF**	**SVC**	**KNN**	**GBC**
Referenzmodell	**DT**		x^2=42,00 p<0,01	x^2=46,68 p<0,01	x^2=26,41 p<0,01	x^2=42,99 p<0,01
	RF	x^2=42,00 p<0,01		x^2=1,70 *p*=0,19	x^2=1,70 *p*=0,19	x^2=0,34 *p*=0,56
	SVC	x^2=46,68 p<0,01	x^2=1,70 *p*=0,19		x^2=5,72 *p*=0,02	x^2=0,50 *p*=0,48
	KNN	x^2=26,41 p<0,01	x^2=1,70 *p*=0,19	x^2=5,72 *p*=0,02		x^2=3,11 *p*=0,08
	GBC	x^2=42,99 p<0,01	x^2=0,34 *p*=0,56	x^2=0,50 *p*=0,48	x^2=3,11 *p*=0,08	

1.3.3 Analyse der Qualitätsparameter

Insgesamt wurden 1.634 EEGs analysiert, und die Krampfqualitätsmetriken sowohl durch das Framework als auch durch das Stimulationsgerät berechnet (siehe Tabelle 5). Zusätzlich wurden 305 EEGs von einem EKT-Experten hinsichtlich Krampfdauer und des PSI bewertet.

- Die Berechnung der Krampfdauer zeigte eine sehr hohe Korrelation zwischen dem Framework und der Expertenbewertung ($\varrho = 0{,}95$, $p < 0{,}01$) sowie zwischen dem Framework und dem Stimulationsgerät ($\varrho = 0{,}95$, $p < 0{,}01$; siehe Abbildung 6 A bis D). Das Histogramm der Krampfdauern zeigte eine symmetrische Verteilung, wobei die Werte des Frameworks, des Stimulationsgeräts und der Expertenbewertung eng beieinanderlagen. Die mittleren Abweichungen betrugen 0,23 Sekunden zwischen Framework und Expertenbewertung ($U = 36.543$, $p = 0{,}807$) sowie 0,28 Sekunden zwischen Framework und Stimulationsgerät ($U = 617.693$, $p = 0{,}798$).

- Die PSI-Werte des Frameworks korrelierten stark mit denen des Stimulationsgeräts ($\varrho = 0{,}94$, $p < 0{,}01$) und mit der Expertenbewertung ($\varrho = 0{,}52$, $p < 0{,}01$; siehe Abbildung 7 A bis D). Die Histogramm-Daten zeigten eine größere Streuung zwischen Expertenbewertung und den automatisierten Methoden, mit einer mittleren Differenz von 14,94 % zwischen Framework und Expertenbewertung ($U = 44.580$, $p < 0{,}01$) sowie 15,72 % zwischen Stimulationsgerät und Expertenbewertung ($U = 45.850$, $p < 0{,}01$). Das Framework und das Stimulationsgerät berechneten jeweils einen mittleren PSI-Wert von 71 % ($U = 584.356$, $p = 0{,}76$).

Tab. 5: Deskriptive Statistiken und Vergleich der Qualitätsparameter zwischen Framework, Stimulationsgerät und Expertenbewertung (modifiziert nach Kayser et al., 2025). Dargestellt sind für jede Teilmenge die Stichprobengröße n, die Anzahl fehlender Werte $\sum_{NA}$, der Mittelwert, die Standardabweichung SD, der Standardfehler des Mittelwerts SEM, der Spearman Rangkorrelationskoeffizient und das Ergebnis des Mann-Whitney-U Tests MWU. Krampfdauer und PSI werden jeweils paarweise zwischen Framework, Stimulationsgerät und Expertenbewertung verglichen. Die weiteren Qualitätsparameter ASEI, EIA, MIA, COH, MSP und TTPP werden zwischen Framework und Stimulationsgerät gegenübergestellt.

		Teilmenge	n	$\sum_{NA}$	Mittelwert	SD	SEM	Spearman	MWU
Qualitätsparameter	Krampfdauer (s)	Framework	272	246	34,44 Δ=0,23	15,59	0,95	ϱ=0,95	U=36543
		Expert		20	34,67	15,71	0,95	p<0,01	p=0,81
		Framework	1108	246	34,15 Δ=0,28	16,19	0,49	ϱ=0,95	U=617693
		Stimulationsgerät		475	33,87	16,21	0,49	p<0,01	p=0,80
		Stimulationsgerät	254	475	34,57 Δ=0,52	15,76	0,99	ϱ=0,95	U=31590
		Expert		20	35,09	15,79	0,99	p<0,01	p=0,69
	PSI (%)	Framework	248	262	69,29 Δ=14,94	22,61	1,44	ϱ=0,52	U=44580
		Expert		25	54,35	18,04	1,15	p<0,01	p<0,01
		Framework	1077	262	71,44 Δ=0,15	21,48	0,65	ϱ=0,94	U=584356
		Stimulationsgerät		524	71,28	21,08	0,64	p<0,01	p=0,76
		Stimulationsgerät	252	524	70,09 Δ=15,72	21,68	1,37	ϱ=0,54	U=45850
		Expert		25	54,37	17,97	1,13	p<0,01	p<0,01
	ASEI (μV^2)	Framework	1354	195	7940,30 Δ=1756,98	9177,77	249,42	ϱ=0,99	U=1037207
		Stimulationsgerät		209	6183,33	7487,66	203,49	p<0,01	p<0,01
	EIA (μV)	Framework	1414	219	68,79 Δ=28,09	32,14	0,85	ϱ=0,91	U=749082
		Stimulationsgerät		74	96,88	61,68	1,64	p<0,01	p<0,01
	MIA (μV)	Framework	1304	253	66,82 Δ=54,42	39,31	1,09	ϱ=0,99	U=428448
		Stimulationsgerät		236	121,24	69,39	1,92	p<0,01	p<0,01
	COH (%)	Framework	1298	219	79,58 Δ=8,09	16,77	0,47	ϱ=0,86	U=545874
		Stimulationsgerät		242	87,67	13,67	0,38	p<0,01	p<0,01
	MSP $\left(\frac{\mu V^2}{Hz}\right)$	Framework	1294	265	13302,21 Δ=3383,23	14690,34	408,38	ϱ=0,99	U=948803
		Stimulationsgerät		234	9918,98	10973,98	305,07	p<0,01	p<0,01
	TTPP (s)	Framework	1307	219	15,41 Δ=1,25	8,56	0,24	ϱ=0,97	U=949914
		Stimulationsgerät		232	14,16	8,56	0,24	p<0,01	p<0,01

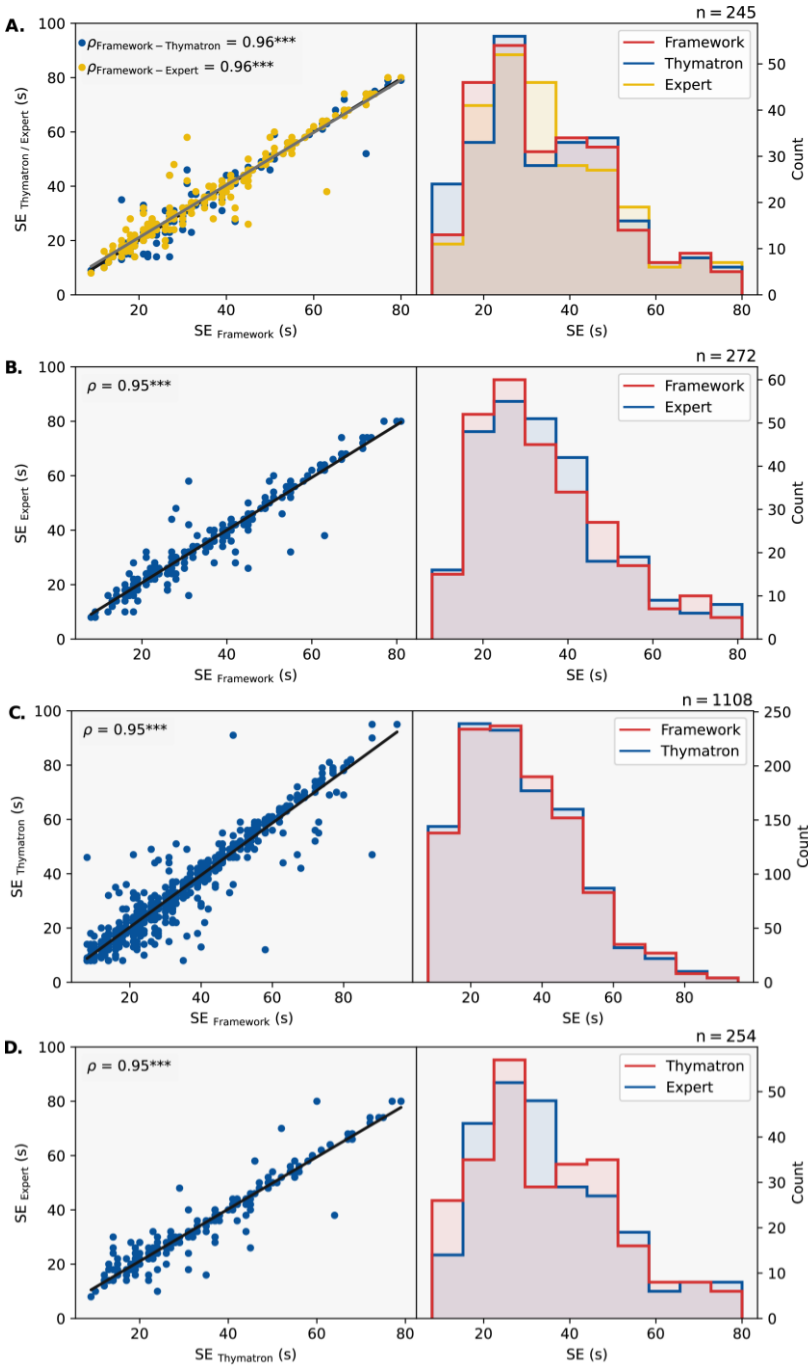

Abb. 6: Spearman Korrelationen und Verteilungen der Krampfdauer für Framework, Stimulationsgerät und Expertenbewertung (modifiziert nach Kayser et al., 2025). Die linken Teilpanels zeigen Streudiagramme der Krampfdauer **(SE) mit Regressionsgeraden und Spearman Rangkorrelationskoeffizienten, die rechten Teilpanels die entsprechenden Histogramme der Verteilungen. In Panel A ist die Beziehung zwischen Framework und Stimulationsgerät (Thymatron®) dargestellt, in Panel B die Beziehung zwischen Framework und Experte. Panel C zeigt die Beziehung zwischen dem Experten und dem Stimulationsgerät, Panel D die Verteilung für Framework, Stimulationsgerät und Experte.**

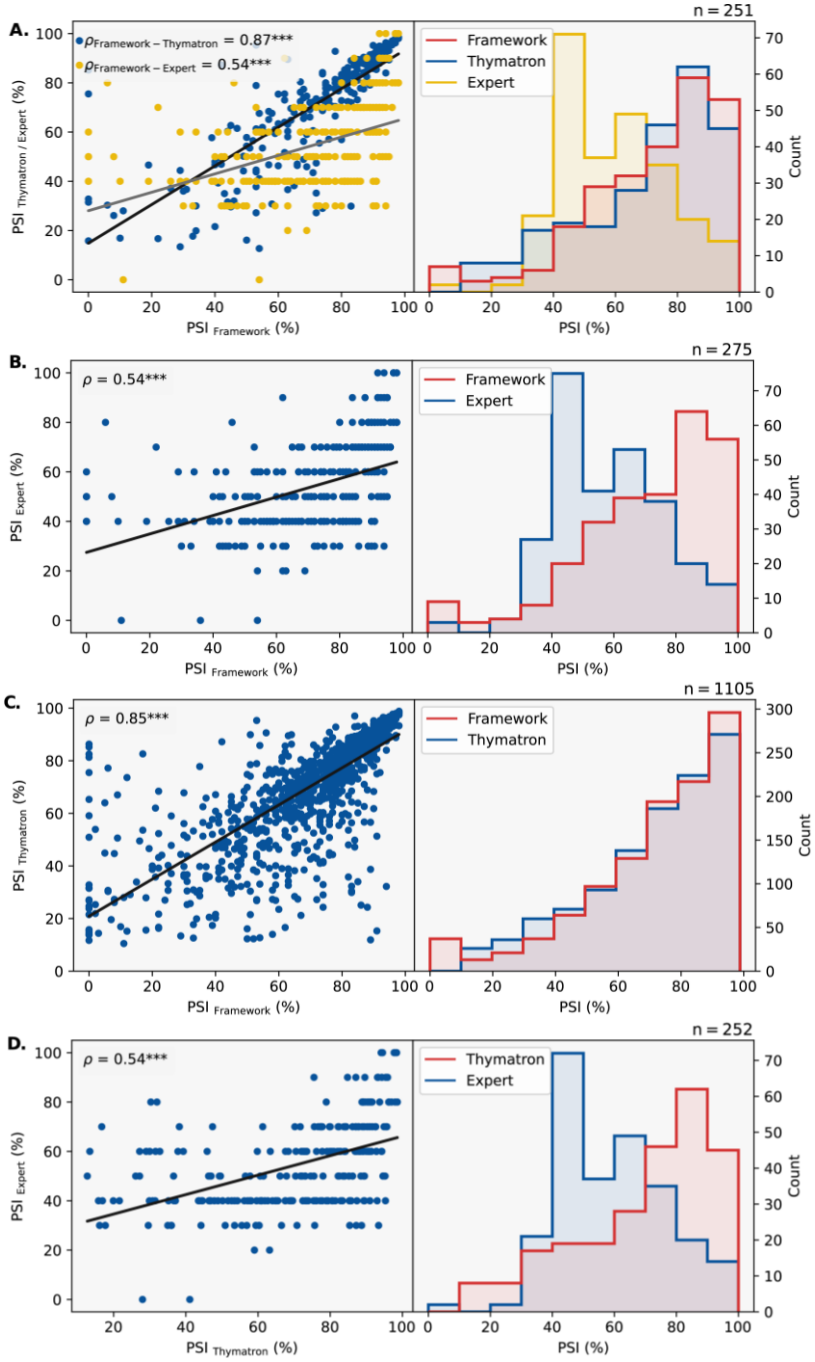

Abb. 7: Spearman Korrelationen und Verteilungen des PSI für Framework, Stimulationsgerät und Expertenbewertung (modifiziert nach Kayser et al., 2025). Die linken Teilpanels zeigen Streudiagramme des PSI mit Regressionsgeraden und Spearman Rangkorrelationskoeffizienten, die rechten Teilpanels die entsprechenden Histogramme der Verteilungen. In **Panel A ist die Beziehung zwischen Framework und Stimulationsgerät (Thymatron®) dargestellt, in Panel B die Beziehung zwischen Framework und Experte. Panel C zeigt die Beziehung zwischen Experten und Stimulationsgerät, Panel D die Verteilung für Framework, Stimulationsgerät und Experte.**

- Die Berechnung des ASEI zeigte eine sehr hohe Korrelation zwischen dem Framework und dem Stimulationsgerät (ϱ = 0,99, p < 0,01; siehe Abbildung 8). Das Histogramm zeigte eine leicht rechts-schiefe Verteilung mit einer mittleren Differenz von 1756,98 µV², die signifikant war (p < 0,05).

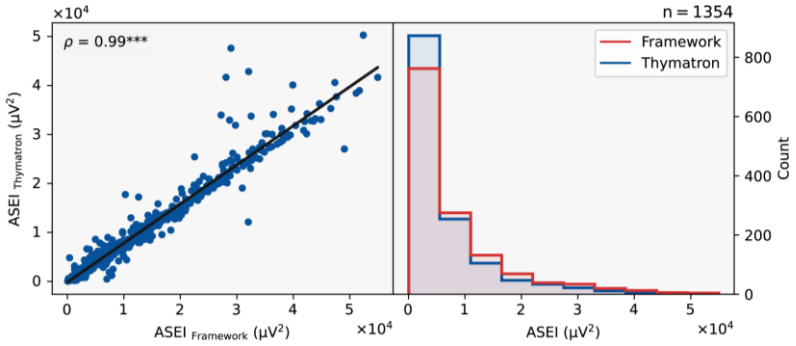

Abb. 8: Spearman Korrelation und Verteilung von ASEI (modifiziert nach Kayser et al., 2025). Spearman Korrelation von ASEI zwischen Framework und **Stimulationsgerät (Thymatron®) links als Streudiagramm mit Regressionsgerade und Rangkorrelationskoeffizient ρ, rechts die Histogramme der Verteilungen für Framework und Stimulationsgerät.**

- Die Korrelation der EIA-Werte zwischen Framework und Stimulationsgerät betrug ϱ = 0,91, p < 0,01 (siehe Abbildung 9). Die Histogramm-Daten wiesen eine größere Streuung im unteren Amplitudenbereich auf, mit einer mittleren Differenz von 28,09 µV, die signifikant war (p < 0,05).

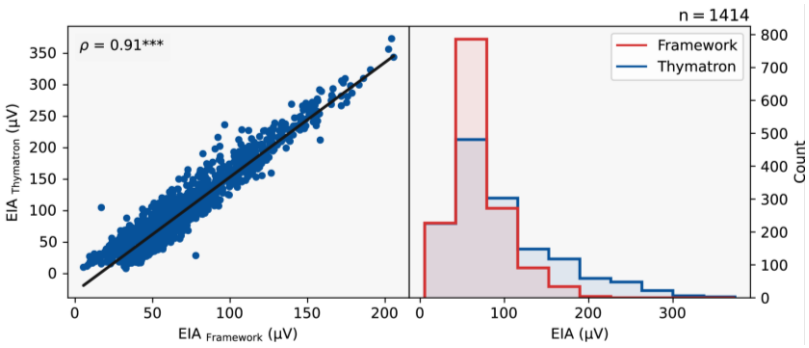

Abb. 9: Spearman Korrelation und Verteilung von EIA (modifiziert nach Kayser et al., 2025). Spearman Korrelation von EIA zwischen Framework und **Stimulationsgerät (Thymatron®) links als Streudiagramm mit Regressionsgerade und Rangkorrelationskoeffizient ρ, rechts die Histogramme der Verteilungen für Framework und Stimulationsgerät.**

- Die Berechnung der MIA ergab eine sehr hohe Korrelation zwischen Framework und Stimulationsgerät (ϱ = 0,99, p < 0,01; siehe Abbildung 10). Das Histogramm zeigte eine breite Werteverteilung, wobei das Framework im Mittel leicht höhere Spitzenwerte berechnete. Die mittlere Differenz betrug 54,42 µV und war signifikant (p < 0,05).

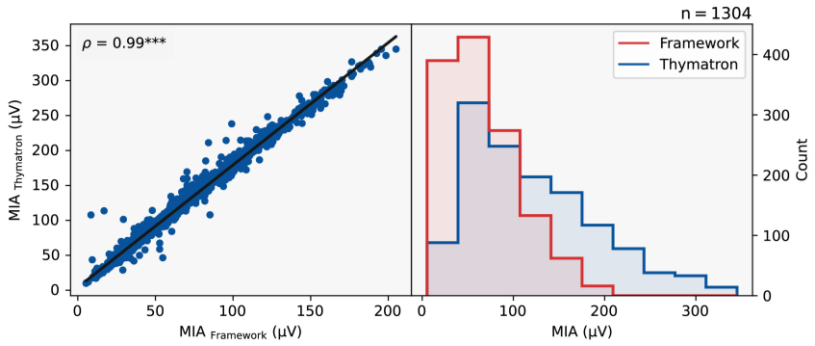

Abb. 10: Spearman Korrelation und Verteilung von MIA (modifiziert nach Kayser et al., 2025). Spearman Korrelation von MIA zwischen Framework und **Stimulationsgerät (Thymatron®) links als Streudiagramm mit Regressionsgerade und Rangkorrelationskoeffizient ρ, rechts die Histogramme der Verteilungen für Framework und Stimulationsgerät.**

- Die COH-Werte korrelierten mit ϱ = 0,86 (p < 0,01) zwischen den beiden Methoden (siehe Abbildung 11). Die Histogramm-Daten zeigten eine größere Varianz in der Kohärenzberechnung, mit einer mittleren Differenz von 8,09 %, die signifikant war (p < 0,05).

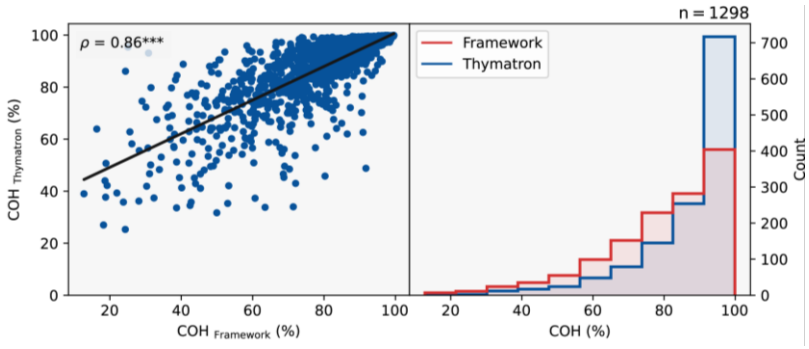

Abb. 11: Spearman Korrelation und Verteilung von COH (modifiziert nach Kayser et al., 2025). Spearman Korrelation von COH zwischen Framework und **Stimulationsgerät (Thymatron®) links als Streudiagramm mit Regressionsgerade und Rangkorrelationskoeffizient ρ, rechts die Histogramme der Verteilungen für Framework und Stimulationsgerät.**

- Die MSP-Werte zeigten eine sehr hohe Korrelation zwischen den Methoden (ϱ = 0,99, p < 0,01; siehe Abbildung 12). Die Histogramm-Daten wiesen auf eine leicht links-schiefe Verteilung hin, mit einer mittleren Differenz von 3383,23 µV²/Hz, die signifikant war (p < 0,05).

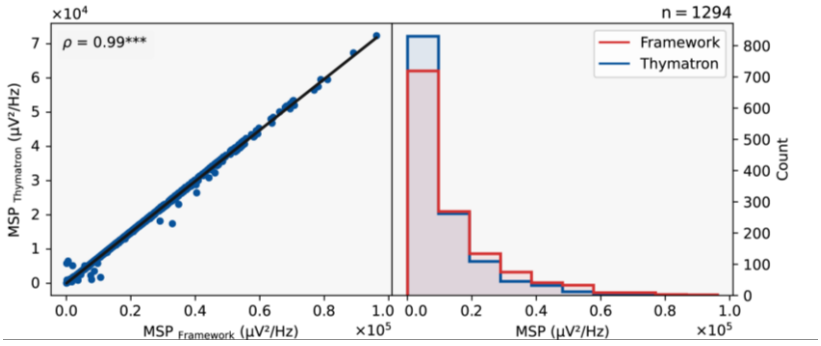

Abb. 12: Spearman Korrelation und Verteilung von MSP (modifiziert nach Kayser et al., 2025). **Spearman Korrelation von MSP zwischen Framework und Stimulationsgerät (Thymatron®) links als Streudiagramm mit Regressionsgerade und Rangkorrelationskoeffizient ρ, rechts die Histogramme der Verteilungen für Framework und Stimulationsgerät.**

- Für die TTPP wurde eine starke Korrelation zwischen Framework und Stimulationsgerät gemessen (ϱ = 0,97, p < 0,01; siehe Abbildung 13). Die Histogramm-Daten zeigten eine leicht breitere Verteilung für das Framework im Vergleich zum Stimulationsgerät, wobei die mittlere Differenz 1.25 Sekunden betrug (p < 0,05).

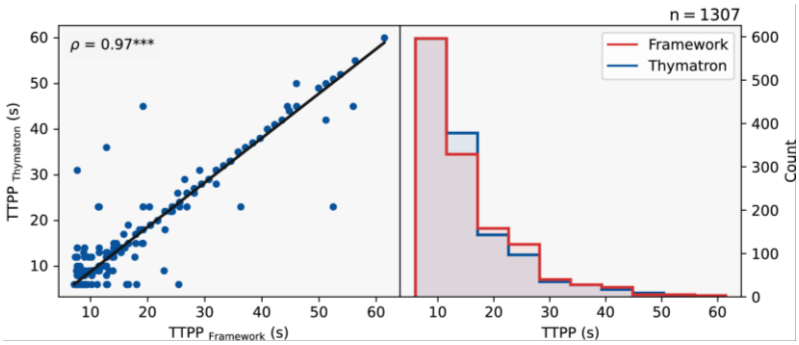

Abb. 13: Spearman Korrelation und Verteilung von TTPP (modifiziert nach Kayser et al., 2025). Spearman Korrelation von TTPP zwischen Framework und **Stimulationsgerät (Thymatron®) links als Streudiagramm mit Regressionsgerade und Rangkorrelationskoeffizient ρ, rechts die Histogramme der Verteilungen für Framework und Stimulationsgerät.**

Der Levene-Test zeigte signifikante Varianzunterschiede für PSI zwischen den Messungen des Frameworks und der Experten sowie zwischen dem Stimulationsgerät und der Expertenbewertung. Zudem wurden signifikante Varianzunterschiede zwischen Framework und Stimulationsgerät für ASEI, EIA, MIA, COH und MSP festgestellt (alle p < 0,01). Der Shapiro-Test ergab, dass keine der berechneten und beobachteten Krampfqualitätsmetriken einer Normalverteilung folgte (alle p < 0,01).

Tab. 6: Intraklassenkorrelationskoeffizienten der Qualitätsparameter (modifiziert nach Kayser et al., 2025). Angegeben sind der Intraklassenkorrelationskoeffizient ICC der Typen ICC3 und ICC3K mit 95 %-Konfidenzintervall sowie die Ergebnisse des zugehörigen F-Tests mit Freiheitsgraden df_1 und df_2 und den jeweiligen p-Werten.

					F-test			
		Typ	ICC	95 % CI	F	df_1	df_2	p-Wert
Qualitätsparameter	Krampfdauer (s)	ICC3	0,97	0,96 - 0,97	87,58	244	488	p<0,01
		ICC3K	0,99	0,99 - 0,99	87,58	244	488	p<0,01
	PSI (%)	ICC3	0,58	0,51 - 0,64	5,16	250	500	p<0,01
		ICC3K	0,81	0,76 - 0,84	5,16	250	500	p<0,01
	ASEI (μV^2)	ICC3	0,96	0,95 - 0,96	47,08	1361	1361	p<0,01
		ICC3K	0,98	0,98 - 0,98	47,08	1361	1361	p<0,01
	EIA (μV)	ICC3	0,78	0,76 - 0,80	8,03	1421	1421	p<0,01
		ICC3K	0,88	0,86 - 0,89	8,03	1421	1421	p<0,01
	MIA (μV)	ICC3	0,85	0,84 - 0,87	12,44	1309	1309	p<0,01
		ICC3K	0,92	0,91 - 0,93	12,44	1309	1309	p<0,01
	COH (%)	ICC3	0,77	0,75 - 0,79	7,76	1303	1303	p<0,01
		ICC3K	0,87	0,86 - 0,88	7,76	1303	1303	p<0,01
	MSP $\left(\frac{\mu V^2}{Hz}\right)$	ICC3	0,96	0,95 - 0,96	46,56	1299	1299	p<0,01
		ICC3K	0,98	0,98 - 0,98	46,56	1299	1299	p<0,01
	TTPP (s)	ICC3	0,88	0,87 - 0,89	15,98	1312	1312	p<0,01
		ICC3K	0,94	0,93 - 0,94	15,98	1312	1312	p<0,01

Die in Tabelle 6 dargestellte Analyse der Intraklassen-Korrelationskoeffizienten zeigte eine exzellente Übereinstimmung der Krampfdauer zwischen Framework, Stimulationsgerät und Experten mit einem ICC_3 von 0,97 (95 % CI: 0,96–0,97) und einem ICC_3K von 0,99 (95 % CI: 0,99–0,99). Für den PSI wurde eine moderate Übereinstimmung zwischen den individuellen Bewertern beobachtet mit einem ICC_3 von 0,58 (95 % CI: 0,51–0,64), während gemittelte Bewertungen eine gute Reliabilität mit einem ICC_3K von 0,81 (95 % CI: 0,76–0,84) zeigten. Die Übereinstimmung zwischen Framework und Stimulationsgerät war für alle weiteren Qualitätsparameter hoch, mit ICC_3-Werten zwischen 0,77 und 0,96. Die niedrigste Übereinstimmung wurde für COH mit einem ICC_3 von 0,77 (95 % CI: 0,75–0,79) gemessen, während die höchste Übereinstimmung für MSP mit einem ICC_3 von 0,96 (95 % CI: 0,95–0,96) bestand. Durchschnittliche Bewertungen verbesserten die Reliabilität weiter, mit ICC_3K-Werten zwischen 0,87 und 0,98. Besonders hohe ICC_3K-Werte wurden für MSP und ASEI mit jeweils 0,98 (95 % CI: 0,98–0,98) sowie für TTPP mit 0,94

(95 % CI: 0,93–0,94) festgestellt. EIA und MIA zeigten ebenfalls eine hohe Übereinstimmung mit ICC_3K-Werten von 0,88 (95 % CI: 0,86–0,89) bzw. 0,92 (95 % CI: 0,91–0,93).

1.4 Diskussion

Im Folgenden werden die Ergebnisse im Lichte der eingangs formulierten Forschungsfragen eingeordnet, ihre methodischen und klinischen Implikationen diskutiert und zentrale Limitationen herausgearbeitet.

1.4.1 Zentrale Befunde

Die Ergebnisse dieser Arbeit zeigen, dass EKT induzierte Krampfanfälle mithilfe einer segmentbasierten Auswertung zweikanaliger EEG-Signale zuverlässig als iktale gegenüber nicht-iktalen Abschnitten klassifiziert und auf dieser Grundlage ein einheitlich definierter Anfallsendpunkt bestimmt werden können. Darauf aufbauend lassen sich etablierte Qualitätsparameter, insbesondere Krampfdauer sowie energie- und amplitudenbasierte Indizes, konsistent berechnen und verlaufsbezogen auswerten. Die Kombination aus standardisierter Datenerfassung, regelgeleiteter Signalverarbeitung und hyperparametrisch optimierter Klassifikation schafft damit einen reproduzierbaren Bezugsrahmen für die digitale Bewertung der Krampfqualität.

1.4.2 Interpretation der Ergebnisse

Die Befunde dieser Arbeit entstehen aus Routinedaten psychiatrischer Universitätskliniken, aus einer strukturierten digitalen Erfassung der EKT und aus einer mehrstufigen Auswertung der EEG-Signale. Das Zusammenspiel von Versorgungsrealität, standardisierter Dateninfrastruktur und methodisch kontrollierter Analysepipeline ist entscheidend für das Verständnis der Ergebnisse. Die folgenden Unterabschnitte betrachten zentrale Aspekte dieser Konstellation und ordnen ihre Bedeutung für die Bewertung und Weiterentwicklung der EKT ein.

1.4.2.1 Datensatz und klinische Charakteristika

Die Befunde beruhen auf klinischen Routinedaten aus zwei deutschen universitären Behandlungszentren der Maximalversorgung. Beide Standorte verfügen über eine vergleichbare Infrastruktur und Prozessqualität. Es wurde dasselbe Stimulationssystem eingesetzt und die Datenerhebung erfolgte einheitlich digital mithilfe von GENET-GPD. Die Behandlungsserien wurden nach ähnlichen klinischen Standards durchgeführt. Unterschiede im Schwellenfindungsverfahren sind dokumentiert, bewegen sich jedoch innerhalb etablierter Vorgehensweisen. Insgesamt liegt damit ein strukturell deckungsgleicher Datensatz vor, der einen gemeinsamen Bezugsrahmen für die Auswertung schafft und die Vergleichbarkeit der Ergebnisse innerhalb des untersuchten Settings unterstützt.

Die Kohorte umfasst erwachsene Patienten mit einem Schwerpunkt im mittleren Erwachsenenalter. Jüngere sowie deutlich ältere Patienten sind weniger häufig vertreten, was die Übertragbarkeit der Ergebnisse auf sehr junge oder hochbetagte Gruppen einschränkt. Der Männeranteil ist leicht erhöht. Klinisch stehen unipolare depressive Episoden im Vordergrund, während bipolare Verläufe zwar vorhanden, aber zahlenmäßig nachgeordnet sind. Die Fallzahl auf Patienten und Behandlungsebene ist für eine kliniknahe retrospektive EKT-Studie angemessen und bildet sowohl interindividuelle Unterschiede als auch serielle Verläufe ab. Die Zahl der Behandlungen pro Serie sowie Stimulationsenergie und Stimulationsdauer bewegen sich überwiegend im üblichen klinischen Bereich. Vereinzelte Randkonstellationen verweisen auf die Heterogenität realer Behandlungsverläufe, ohne den Gesamteindruck einer typischen universitären EKT-Versorgung grundlegend zu verändern. Bei der Elektrodenplatzierung dominieren unilateral rechts und bifrontal, während bitemporale und links anterior rechts temporale Montagen selten sind und unilateral links eine Ausnahme darstellt. Die zunehmende Verbreitung dieser Standardmontagen unterstützt die Relevanz der Befunde für vergleichbare Versorgungssettings, lässt aber offen, in welchem Ausmaß sehr seltene Montagen und Sonderkonstellationen erfasst werden.

1.4.2.2 Implementierung und Reproduzierbarkeit

Die Auswertung beruht auf einer vollständig skriptbasierten Implementierung in einer einheitlichen Programmierumgebung. Für alle Schritte der Signalverarbeitung kamen etablierte und frei verfügbare Softwarebibliotheken zum Einsatz, unter anderem für Vorverarbeitung, Merkmalsextraktion und Klassifikation. Eine detaillierte Übersicht aller verwendeten Softwarebibliotheken einschließlich Version, Autor und Verwendungszweck findet sich in Supplement A der Originalpublikation (Kayser et al., 2025). Die in Abbildung 1 dargestellte Abfolge der Verarbeitungsschritte wurde in dieser Umgebung vollständig umgesetzt und unverändert auf alle Datensätze angewandt. Dadurch ist das Analyse-Framework von der Rohdatenebene bis zu den abgeleiteten Qualitätsparameter klar strukturiert und vollständig nachvollziehbar.

Durch die Festlegung eines einheitlichen zufälligen Startwertes und die Verwendung dokumentierter Versionen der verwendeten Bibliotheken wird der Einfluss zufälliger Effekte und technischer Änderungen begrenzt. Dies erleichtert interne Plausibilitätsprüfungen und unterstützt Replikationsversuche in anderen Arbeitsgruppen. Die Nachvollziehbarkeit der Ergebnisse stützt sich damit nicht allein auf verbale Methodenbeschreibungen, sondern auf konkret ausführbaren Code und folgt den Prinzipien reproduzierbarer Forschung.

Mit der `ECTMetrics` Toolbox steht eine Implementierung zur Verfügung, welche die Berechnung der verwendeten Qualitätsparameter in standardisierter Form ermöglicht (Kayser, 2025). Kliniken und Forschungseinrichtungen können eigene Datensätze mit denselben Algorithmen auswerten und ihre Ergebnisse unmittelbar mit den hier berichteten Befunden vergleichen. Damit entsteht die Möglichkeit, über mehrere Zentren hinweg eine gemeinsame digitale Referenz für die Bewertung der Krampfqualität aufzubauen. Zugleich macht die offene Implementierung die zugrunde liegenden Annahmen transparent und reduziert die Abhängigkeit von geräteinternen, nicht einsehbaren Berechnungsroutinen. Das fördert eine wissenschaftlich reflektierte Einordnung und Weiterentwicklung der verwendeten Kennwerte durch die Fachgemeinschaft.

1.4.2.3 Stichprobentiefe und Auswertungsschnitt

Die segmentbasierte Auswertung erhöht die Zahl der verfügbaren Beobachtungen deutlich und bildet die zeitliche Struktur des Krampfgeschehens differenziert ab. Innerhalb einer Behandlung fließen zahlreiche Segmente in die Analyse ein, die unterschiedliche Phasen des iktalen und nicht-iktalen Verlaufs widerspiegeln. Aus methodischer Sicht entsteht dadurch ein großes nominales Stichprobenvolumen, das die Modelle wiederholt mit ähnlichen, aber nicht identischen Mustern konfrontiert und so ein stabileres Lernen der relevanten Signalcharakteristika ermöglicht.

Gleichzeitig stammen viele Segmente aus denselben Behandlungen und beziehen sich auf dieselben Patienten. Die Modelle sehen charakteristische Muster einzelner Behandlungen daher wiederholt. Für die Interpretation bedeutet dies, dass die nominale Segmentzahl nicht mit einer entsprechend großen Zahl unabhängiger Fälle gleichgesetzt werden darf. Die Auswertung erlaubt eine praxisnahe Einschätzung der Leistungsfähigkeit des Frameworks innerhalb des vorliegenden Versorgungsrahmens, lässt aber bewusst offen, wie sich die Leistung in deutlich abweichenden klinischen Konstellationen darstellt.

Die gewählte Auswertung spiegelt die Versorgungsrealität der beteiligten Zentren wider und erlaubt eine praxisnahe Einschätzung der Leistungsfähigkeit der Pipeline in einem vergleichbaren Umfeld. Sie zeigt, dass sich unter typischen Bedingungen der Routineversorgung charakteristische iktale Muster zuverlässig erkennen lassen. Aussagen zur Übertragbarkeit auf bislang nicht beobachtete Konstellationen etwa andere Geräte, stark abweichende Stimulationsstrategien oder sehr spezielle Patientengruppen bleiben damit jedoch begrenzt. Eine weiterführende Validierung mit strikt patientenbasierten Auswertungsstrategien und mit externen Datensätzen ist notwendig, um die Generalisierbarkeit der Befunde über das hier untersuchte Setting hinaus zu sichern.

1.4.2.4 Merkmalsraum

Der in dieser Arbeit verwendete Merkmalsraum bildet verschiedene physiologisch und technisch relevante Aspekte des EKT induzierten Krampfgeschehens ab. Energie- und amplitudenbasierte Kennwerte, spektrale Merkmale und Maße der zeitlichen Struktur tragen jeweils unterschiedliche Informationen bei. Einige Variablen beschreiben vor allem

Stärke und Ausprägung der Aktivität, andere die Verteilung über Frequenzbereiche oder den zeitlichen Verlauf. Erst ihre gemeinsame Betrachtung ermöglicht eine hinreichend differenzierte Charakterisierung von iktalen und nicht-iktalen Muster.

Die Merkmalsauswahl orientiert sich daran, sowohl etablierte EEG-Parameter als auch aus der klinischen Praxis bekannte Einschätzungsdimensionen abzubilden. Damit wird die Brücke zwischen traditioneller visueller Beurteilung und automatisierter numerischer Auswertung geschlagen. Merkmale, die sich als redundant oder nur gering informativ erwiesen haben, wurden in der weiteren Auswertung zurückgestellt. Auf diese Weise entstand ein kompakterer Merkmalsraum, der die wesentlichen Unterschiede zwischen iktaler und nicht-iktaler Aktivität erfasst, ohne die Modelle mit unnötig vielen, stark korrelierten Variablen zu belasten.

Die anschließende Verdichtung auf wenige, aus den verbleibenden Variablen abgeleitete Komponenten zielt auf ein Gleichgewicht zwischen Informationsgehalt und Modellkomplexität. Die Klassifikatoren operieren damit in einem Raum, der wichtige Strukturunterschiede bewahrt und gleichzeitig das Risiko von Überanpassung reduziert. Zugleich bleiben die zugrunde liegenden Kennwerte inhaltlich interpretierbar, da sie unmittelbar auf bekannte EEG-Merkmalsfamilien zurückgeführt werden können. Im Vergleich zu Ansätzen, die sich überwiegend auf einzelne Parameter wie Krampfdauer oder grobe Frequenzanteile stützen, ermöglicht dieser Merkmalsraum eine systematischere und besser standardisierbare Beschreibung der Krampfqualität.

1.4.2.5 Hyperparameteroptimierung

Die Einstellung der Hyperparameter zielte darauf, eine möglichst ausgewogene Leistungsfähigkeit bei ungleichen Klassenhäufigkeiten zu erreichen. Im Vordergrund stand die zuverlässige Unterscheidung von iktalen und nicht-iktalen Segmenten in beiden Richtungen, also das Erkennen von Anfällen und das Vermeiden unnötiger Fehlalarme. Unter diesen Bedingungen ist eine Konfiguration erforderlich, die weder zu stark vereinfacht noch unnötig komplex ist. Zu einfache Modelle würden relevante Strukturunterschiede nicht ausreichend abbilden, während übermäßig flexible Modelle dazu tendieren, zufällige Muster im Rauschen der Daten zu lernen.

Auffällig ist, dass unterschiedliche Modellfamilien nach der Optimierung auf ein ähnliches Leistungsniveau gelangen. Dies spricht dafür, dass die entscheidenden Informationen zur Trennung iktaler und nicht-iktaler Aktivität bereits durch Signalaufbereitung und Merkmalsrepräsentation erschlossen werden. Die konkrete Wahl des Klassifikationsverfahrens tritt damit hinter die Qualität der Vorverarbeitung und die Struktur des Merkmalsraums zurück. Für zukünftige Arbeiten deutet dies darauf hin, dass methodische Fortschritte weniger durch einen Wechsel der Modellarchitektur als durch eine Verfeinerung der Merkmalsauswahl, der Referenzdefinitionen und der Einbindung zusätzlicher Datenquellen zu erwarten sind.

Gleichzeitig zeigen die Ergebnisse der Hyperparameteroptimierung, dass sich die Modelle in einem Bereich mittlerer Komplexität am stabilsten verhalten. Konfigurationen mit sehr hoher Flexibilität bringen keinen erkennbaren Gewinn, erhöhen aber das Risiko von Überanpassung und erschweren die Interpretierbarkeit. Die hier gefundene Balance unterstützt daher nicht nur eine robuste Leistung, sondern bleibt auch mit klinischen Anforderungen vereinbar, in denen Nachvollziehbarkeit und technische Stabilität eine zentrale Rolle spielen.

1.4.2.6 Leistungsbewertung der Modelle

Die untersuchten Modelle liefern insgesamt eine stabile und klinisch gut nutzbare Trennung zwischen iktaler und nicht-iktaler Aktivität auf Segmentebene. Unterschiede zwischen den leistungsstärkeren Ansätzen zeigen sich vor allem in Details der Fehlerverteilung und im Umgang mit Grenzbereichen, nicht in grundlegend verschiedenen Leistungsniveaus. Ein einzelnes überragendes Verfahren lässt sich unter den hier betrachteten Bedingungen nicht erkennen. Vielmehr bestätigt sich, dass mehrere Modellklassen in der Lage sind, die im Merkmalsraum enthaltenen Informationen in vergleichbarer Weise auszuschöpfen.

Für die praktische Anwendung rückt damit die Qualität der gesamten Auswertungskette in den Vordergrund. Eine sorgfältige Aufbereitung der EEG-Signale, eine inhaltlich begründete Auswahl und Verdichtung der Merkmale und eine konsistente Referenzdefinition sind mindestens so bedeutsam wie die konkrete Wahl des Klassifikators. Der Random

Forest bietet sich in diesem Rahmen als pragmatisches Referenzmodell an, da er eine günstige Verbindung aus Robustheit gegenüber Ausreißern, angemessener Komplexität und nachvollziehbarem Entscheidungsverhalten aufweist. Gleichzeitig bleibt zu berücksichtigen, dass auch gut bewertete Modelle in Randbereichen Fehlentscheidungen treffen. Besonders von Artefakten stark beeinflusste Abschnitte, sehr kurze oder ungewöhnliche Krampfverläufe und seltene Konstellationen der Stimulations- und Patientencharakteristika erfordern weiterhin eine kritische fachärztliche Beurteilung und dürfen nicht ausschließlich automatisierten Entscheidungen überlassen werden.

1.4.2.7 Qualitätsparameter

Die aus dem automatisiert bestimmten Anfallsendpunkt abgeleiteten Qualitätsparameter stellen die direkte Verbindung zwischen der technischen Auswertung und der klinischen Beurteilung der EKT dar. Besonders Krampfdauer sowie energie- und amplitudenbasierte Indizes erweisen sich als tragfähige Kernparameter. Sie bilden die Ausprägung des iktalen Geschehens konsistent ab und zeigen eine hohe strukturelle Übereinstimmung mit den geräteinternen Berechnungen und den Bewertungen durch erfahrene Behandler. Diese Metriken eignen sich deshalb gut für eine standardisierte, digital gestützte Dokumentation der Krampfqualität und erlauben eine nachvollziehbare Beschreibung von Verlaufsänderungen innerhalb einer Serie sowie Vergleiche zwischen Behandlungen und Patienten.

Gleichzeitig treten Begrenzungen einzelner traditionell verwendeter Kennwerte deutlich hervor. Der PSI ist in besonderem Maß von der gewählten Definition des Anfallsendes, von der verwendeten Skala und von subjektiven Einschätzungen abhängig. Diskrete Bewertungsstufen und Unterschiede in der visuellen Interpretation führen zu Abweichungen zwischen automatischer Berechnung, geräteinternen Angaben und Expertenurteilen. Der Index spiegelt damit nicht nur die physiologische Erholung nach dem iktalen Ereignis wider, sondern auch Beobachter- und Definitionsvariabilität. Dies schränkt seine Eignung als alleiniger Referenzparameter für eine geräteübergreifende Standardisierung ein.

Die in den Ergebnissen beschriebenen systematischen Niveauunterschiede zwischen Framework, Stimulationsgerät und Expertenbewertung bei energie- und amplitudenbasierten Kennwerten spiegeln vor allem definitorische und algorithmische Besonderheiten wider (z.B. Wahl des Anfallsendpunkts und der Auswertefenster). Trotz dieser Verschiebungen bleibt die Rangordnung der Fälle über die Methoden hinweg weitgehend stabil, was die Eignung dieser Indizes als robuste Kernparameter einer digitalen Krampfqualitätsbewertung stützt.

Kohärenzbasierte Kennwerte reagieren zusätzlich empfindlich auf Elektrodenmontage, Filtereinstellungen, Störquellen und auf Details der Signalaufbereitung. Schon moderate Unterschiede in diesen Rahmenbedingungen können zu relevanten Abweichungen in den resultierenden Werten führen, obwohl das zugrunde liegende iktale Geschehen klinisch vergleichbar ist. Damit besteht die Gefahr, dass Variation im technischen Setup stärker abgebildet wird als tatsächliche Unterschiede in der Krampfqualität. Solche Metriken eignen sich deshalb vor allem für streng standardisierte Protokolle und für Fragestellungen, in denen gerade diese Sensitivität gezielt genutzt werden soll, etwa in experimentellen Studien mit eng geführten Bedingungen.

In der Gesamtschau zeigt sich, dass nicht alle verfügbaren Qualitätsparameter in gleicher Weise für eine breite, geräteunabhängige Standardisierung geeignet sind. Krampfdauer und energie- beziehungsweise amplitudenbasierte Indizes erscheinen als stabiles Fundament einer routinetauglichen digitalen Krampfqualitätsbewertung. Sie verbinden eine hohe Robustheit gegenüber technischen und beobachterbezogenen Einflüssen mit einer unmittelbaren klinischen Deutbarkeit. Der postiktale Suppressionsindex und kohärenzbasierte Maße behalten ihren Wert als ergänzende Informationsquellen. Sie sollten jedoch vor allem unter klar definierten Rahmenbedingungen und im Bewusstsein ihrer Grenzen eingesetzt werden. Eine bewusste Auswahl und Kombination der Metriken ist damit zentral, um eine Beurteilung der Krampfqualität zu erreichen, die sowohl klinisch aussagekräftig als auch methodisch verlässlich ist und die Besonderheiten iktaler und nicht-iktaler Aktivität angemessen berücksichtigt.

1.4.3 Limitationen

Die Aussagekraft der vorliegenden Befunde wird durch mehrere methodischen und inhaltlichen Einschränkungen begrenzt. Diese betreffen die Zusammensetzung der untersuchten Kohorte, das technische und organisatorische Setting der EKT, die gewählte Auswertungsstrategie und die verwendeten Referenzdefinitionen. Eine differenzierte Betrachtung dieser Punkte ist notwendig, um Reichweite und Grenzen der Ergebnisse realistisch einzuschätzen und Ansatzpunkte für zukünftige Arbeiten zu identifizieren.

1.4.3.1 Datengrundlage und Kohorten-Struktur

Die Datengrundlage beschränkt sich auf zwei universitäre Zentren mit vergleichbarer Versorgungsstruktur. Dies stärkt die interne Konsistenz der Befunde, begrenzt jedoch die Generalisierbarkeit auf Kliniken mit abweichenden Rahmenbedingungen. Einrichtungen mit anderen Patientenkollektiven, Versorgungsaufträgen oder stark spezialisierten Indikationsspektren sind in der vorliegenden Arbeit nicht abgebildet. Hinzu kommt, dass sehr junge und sehr alte Patienten sowie seltene Indikationen nur in kleiner Zahl vorkommen. Für diese Gruppen kann aus den vorliegenden Ergebnissen deshalb nur eingeschränkt auf die Leistungsfähigkeit des Frameworks und die Aussagekraft der Qualitätsparameter geschlossen werden.

1.4.3.2 Setting und Gerätetechnik

Die vorliegenden Befunde sind an ein spezifisches technisches und organisatorisches Setting gebunden. Beide Zentren nutzen dasselbe Stimulationsgerät mit zweikanaliger EEG-Ableitung und arbeiten mit standardisierten Routinen der Narkoseführung und Stimulationsplanung. Dies schafft vergleichbare Rahmenbedingungen und erleichtert die Zusammenführung der Daten, beschränkt die Ergebnisse jedoch auf Konstellationen, in denen ein ähnliches Zusammenspiel von Gerät, Ableitungsmontage und klinischem Ablauf vorliegt.

Die zweikanalige Ableitung bildet vor allem frontotemporale Regionen ab und orientiert sich an in der klinischen Praxis etablierten Montagen. Eine feinere räumliche Auflösung

kortikaler Prozesse, wie sie mit hochauflösenden multikanaligen EEG-Systemen möglich wäre, steht damit nicht zur Verfügung. Die hier gewonnenen Modelle und Krampfqualitätsmetriken basieren folglich auf einem begrenzten räumlichen Blick auf das iktale Geschehen. In Settings mit deutlich abweichender Elektrodenkonfiguration oder mit anderen Referenztechniken ist nicht gesichert, dass sich dieselben Muster und Trennkriterien in gleicher Weise wiederfinden.

Hinzu kommt, dass viele Schritte der Signalaufbereitung an die Eigenschaften des verwendeten Stimulationssystems angepasst sind. Filtereinstellungen, Artefaktunterdrückung und die Art der Triggerung orientieren sich an den Möglichkeiten und Konventionen dieses Gerätes. In anderen Systemen können bereits kleine Unterschiede in der technischen Implementierung dazu führen, dass Zeitverläufe, Amplituden und spektrale Anteile modifiziert erscheinen, obwohl das zugrunde liegende iktale Geschehen vergleichbar ist. Die Übertragung der Ergebnisse auf andere Geräte erfordert deshalb eine sorgfältige Prüfung und gegebenenfalls eine Anpassung der Vorverarbeitung.

Auch organisatorische Aspekte des Settings wirken als begrenzende Faktoren. Die beteiligten Zentren verfügen über hohe Fallzahlen, spezialisierte Teams und eine etablierte Dokumentationsinfrastruktur. In kleineren Einrichtungen mit geringerer EKT-Frequenz, weniger standardisierten Abläufen oder eingeschränkten Ressourcen für digitale Dokumentation können sich andere Fehlerprofile und Datenqualitäten ergeben. Das hier entwickelte Framework setzt eine bestimmte Mindestqualität der EEG-Aufzeichnung und Metadaten voraus. Wo diese Anforderungen nicht erfüllt werden, ist mit eingeschränkter Anwendbarkeit oder mit systematischen Verzerrungen zu rechnen.

1.4.3.3 Analyseverfahren und Auswertung

Die gewählte Auswertungsstrategie beruht auf einer segmentbasierten Betrachtung der EEG-Signale und auf einer Aufteilung in Trainings- und Testdaten innerhalb des vorhandenen Gesamtdatensatzes. Dies erlaubt eine differenzierte Analyse des zeitlichen Verlaufs der iktalen und nicht-iktalen Aktivität und schafft eine praktikable Struktur für das Training und die Überprüfung der Modelle. Gleichzeitig führt dieser Ansatz zu mehreren

Einschränkungen, die bei der Interpretation der Ergebnisse berücksichtigt werden müssen.

Die segmentbasierte Auswertung geht mit einer eingeschränkten statistischen Unabhängigkeit der Beobachtungen einher. Segmente derselben Behandlung teilen klinische und technische Rahmenbedingungen, sodass die effektive Zahl unabhängiger Einheiten näher an der Patienten- und Sitzungszahl liegt als an der Gesamtheit der Segmente. Damit besteht grundsätzlich das Risiko, dass Modelle Teile der individuellen Muster einzelner Patienten mitlernen und ihre Leistung in der internen Auswertung überschätzt wird. Eine strikt patientenbasierte Trennung von Trainings- und Testdaten würde dieses Risiko verringern, konnte im Rahmen der vorliegenden Arbeit aber nur begrenzt umgesetzt werden und sollte in zukünftigen Studien gezielt geprüft werden.

Für die abschließende Evaluierung der Modellleistung wurde ein klassischer Holdout-Ansatz mit einer festen Aufteilung in Trainings- und Testdaten verwendet. Obwohl die Hyperparameteroptimierung zuvor bereits mittels fünffacher Kreuzvalidierung erfolgte, begrenzt diese Form der Endauswertung die Aussagen zur Stabilität der Ergebnisse gegenüber alternativen Aufteilungen der Daten. Wiederholte Auswertungen mit unterschiedlichen Aufteilungen oder mit verschachtelten Validierungsstrategien könnten zeigen, wie robust die beobachtete Modellleistung gegenüber zufälligen Schwankungen im Datenzuschnitt ist. In der vorliegenden Arbeit war eine solche systematische Variation aufgrund der Komplexität der Implementierung und des Rechenaufwands nur in begrenztem Umfang möglich. Es bleibt daher eine Unsicherheit darüber, wie stark die berichtete Leistung von der konkreten Wahl der Aufteilung abhängt.

Schließlich beeinflusst die Definition der Auswertungsebene, auf der die Leistung bewertet wird, die Interpretation der Ergebnisse. Die hier verwendete segmentbasierte Bewertung bildet vor allem die Fähigkeit ab, kurze Ausschnitte als iktal oder nicht-iktal zu klassifizieren. Klinische Entscheidungen orientieren sich jedoch häufig an Informationen auf Sitzungs- oder Serienebene. Die Übertragung der segmentbasierten Gütemaße auf diese Ebenen ist nicht trivial und hängt von weiteren Annahmen ab, etwa zu Aggregationsregeln und zur Gewichtung einzelner Segmente. Diese Fragen konnten im Rahmen der vorliegenden Arbeit nur begrenzt adressiert werden und stellen einen wichtigen Ansatzpunkt für zukünftige Untersuchungen dar.

1.4.3.4 Referenz- und Definitionsabhängigkeit

Die Modelle und Qualitätsparameter dieser Arbeit stützen sich auf Referenzen, die ihrerseits nicht frei von Unsicherheiten sind. Als Bezugspunkte dienen die geräteinternen Berechnungen des Stimulationssystems und die visuellen Bewertungen erfahrener Behandler. Beide Quellen repräsentieren den aktuellen klinischen Standard, bilden jedoch weder eine absolut fehlerfreie noch eine vollständig einheitliche Grundlage. Unterschiede in der Auslegung von Kriterien, in der Gewichtung einzelner Signalaspekte und in der Aufmerksamkeit für Störquellen können zu systematischen Abweichungen führen, ohne dass sich eindeutig sagen lässt, welche Einschätzung im Einzelfall den „wahren“ Verlauf vollständig trifft.

Besonders sichtbar wird diese Problematik beim postiktalen Suppressionsindex und bei weiteren Kennwerten, die sich stark an visuellen Eindrücken orientieren. Die Festlegung des Anfallsendes, die Einteilung in Stufen und die Bewertung von Übergangsphasen zwischen iktaler und nicht-iktaler Aktivität beruhen auf Konventionen und Erfahrung, nicht auf eindeutig scharf gefassten naturwissenschaftlichen Grenzen. Das Stimulationsgerät setzt diese Kriterien in algorithmische Regeln um, während Experten ihre Bewertungen situativ und mit Blick auf den Gesamtverlauf treffen. Die automatisierte Auswertung orientiert sich an diesen Referenzen und übernimmt damit deren Stärken und Schwächen.

Auch bei scheinbar objektiveren Parametern wie Krampfdauer und energiebezogenen Indizes bleibt die Definitionsfrage relevant. Bereits kleine Verschiebungen der Grenze zwischen iktaler und postiktaler Aktivität können die resultierenden Werte verändern, obwohl sich der klinische Gesamteindruck kaum unterscheidet. Die hier verwendete segmentbasierte Klassifikation erzeugt eine konsistente und reproduzierbare Festlegung des Anfallsendes, ist jedoch an die zugrunde liegenden Trainingsdaten und Entscheidungsregeln gebunden. Ein anderer Algorithmus oder eine alternative Definition des Endpunkts könnte bei gleichen Rohdaten zu etwas abweichenden Werten führen.

Diese Referenz- und Definitionsabhängigkeit begrenzt die Übertragbarkeit der absoluten Werte auf andere Systeme und Bewertungsansätze. Die Stärke der vorliegenden Arbeit liegt vor allem in der internen Konsistenz innerhalb des gewählten Referenzrahmens. Die

Modelle bilden die in diesem Rahmen etablierten Konzepte von Krampfqualität nachvollziehbar und reproduzierbar ab. Für eine geräte- und zentrumsübergreifende Harmonisierung ist es jedoch erforderlich, die zugrunde liegenden Definitionen weiter zu schärfen und idealerweise in multizentrischen prospektiven Vergleichsstudien zu überprüfen.

1.4.4 Beantwortung der Fragestellungen

Die erste Fragestellung zielte auf die Zuverlässigkeit einer maschinellen Bestimmung des Anfallsendpunkts im Vergleich zu den etablierten klinischen Referenzen. Auf Grundlage der vorliegenden Auswertungen lässt sich diese Frage bejahen. Das entwickelte Framework liefert einen Endpunkt, der sich inhaltlich schlüssig in den bestehenden Beurteilungsrahmen einfügt und klinisch nachvollziehbare Unterschiede zwischen Behandlungen sichtbar macht. Abweichungen zwischen automatisierter Bestimmung, geräteinternen Angaben und visueller Bewertung treten vor allem in Übergangsbereichen und bei Definitionsspielräumen auf und verweisen eher auf konzeptionelle Unterschiede als auf technische Unzuverlässigkeit. Insgesamt kann die datengetriebene Endpunktbestimmung als potenzielle Ergänzung der bisherigen klinischen Praxis angesehen werden.

Die zweite Fragestellung untersuchte, ob etablierte Krampfqualitätsmaße auf Grundlage der automatisierten EEG-Auswertung in einer für die klinische Anwendung ausreichend zuverlässigen Qualität abgeleitet werden können. Für Krampfdauer und energie- beziehungsweise amplitudenbasierte Indizes fällt die Bilanz eindeutig positiv aus. Diese Kennwerte lassen sich mit dem vorgestellten Framework so berechnen, dass sie an vertraute Bewertungsmuster anschließen und eine differenzierte Verlaufsbeschreibung ermöglichen. Kritischer ist die Situation bei Parametern, die in hohem Maß von Definitionskonventionen und subjektiven Einschätzungen abhängen, insbesondere beim postiktalen Suppressionsindex und bei kohärenzbasierten Größen. Hier werden Unterschiede zwischen Auswertewegen stärker sichtbar, ohne dass sich eine eindeutige Referenz festlegen lässt. Insgesamt zeigt sich jedoch, dass der Kern der in der Praxis genutzten Qualitätsparameter mit den vorgestellten Verfahren zuverlässig abgebildet werden kann.

Die dritte Fragestellung richtete sich auf das methodische Potenzial des Frameworks über die reine Anfallsdetektion hinaus. Die modulare Struktur, die standardisierte Endpunktdefinition und die Möglichkeit, zusätzliche Merkmalsfamilien und Datenquellen einzubinden, eröffnen hier vielfältige Erweiterungsoptionen. Die vorliegenden Ergebnisse legen nahe, dass das Framework nicht nur für die retrospektive Auswertung einzelner Serien geeignet ist, sondern auch als Rahmen für multizentrische Vergleiche und für die systematische Verknüpfung von Krampfqualitätsprofilen mit klinischen Verläufen dienen kann. Im Sinne dieser dritten Fragestellung liefern die Ergebnisse einen methodischen Proof-of-Concept: Das Framework ermöglicht nicht nur die Rekonstruktion etablierter Krampfqualitätsparameter, sondern schafft mit der offenen Implementierung in Form der Python-Toolbox `ECT-Metrics` eine erweiterbare Grundlage für zukünftige prädiktive Modelle, beispielsweise zur Vorhersage von Therapieansprechen oder kognitiven Nebenwirkungen (Kayser, 2025). Damit wird deutlich, dass die Arbeit über eine reine Machbarkeitsstudie hinausgeht und einen tragfähigen Ausgangspunkt für weiterführende Forschungs- und Entwicklungsarbeiten im Bereich digital gestützter EKT bietet.

1.4.5 Klinische Implikationen

Die vorliegenden Ergebnisse zeigen, dass sich die Krampfqualität im Rahmen der EKT mithilfe eines standardisierten digitalen Frameworks systematisch, transparent und reproduzierbar erfassen lässt. Damit wird die bisher überwiegend geräte- und erfahrungsabhängige Beurteilung des iktalen Geschehens um eine strukturierte, algorithmisch gestützte Ebene ergänzt. Die automatisierte Bestimmung des Anfallsendpunkts und die darauf aufbauende Berechnung von Krampfdauer sowie energie- und amplitudenbasierten Indizes schafft eine nachvollziehbare Datengrundlage, auf der Behandlerinnen und Behandler intraindividuelle Verläufe innerhalb einer Serie verfolgen, Stimulationsparameter gezielt anpassen und ihre Entscheidungen gegenüber Patienten und Konsiliardiensten begründen können.

Ein wesentlicher Vorteil des Frameworks besteht darin, dass es bewusst an bereits etablierte klinische Heuristiken anschließt. Krampfdauer und amplitudenbezogene Qualitätsparameter gehören seit langem zu den zentralen Orientierungspunkten für die Steuerung

der Behandlung. Ihre standardisierte, automatisierte Berechnung reduziert die Anfälligkeit für manuelle Ablesefehler, unterschiedliche Anzeigeformate und situative Einschätzungen, ohne den vertrauten Referenzrahmen der Behandelnden zu verlassen. Die digital berechneten Metriken bilden zentrale Dimensionen des Krampfgeschehens ab und machen diese sowohl über Serien hinweg als auch zwischen unterschiedlichen Behandlern und Zentren vergleichbar und auditierbar. Das Framework zielt damit nicht auf eine Substitution klinischer Erfahrung, sondern auf deren Kalibrierung an konsistente Messgrößen.

Die Befunde mahnen zugleich zu einer differenzierten Nutzung einzelner Kennwerte. Am Beispiel des PSI wird deutlich, dass die Aussagekraft eines Parameters empfindlich von Definition des Anfallsendes, verwendeter Skala und visueller Einschätzung abhängt. Der PSI eignet sich daher nur begrenzt als alleiniger Steuerungsparameter und sollte in der Praxis im Verbund mit robusteren Metriken, wie Krampfdauer und stabilen amplitudenbasierten Indizes, interpretiert werden. Entsprechendes gilt für kohärenzbasierte Maße, die stark von Elektrodenmontage, Filtereinstellungen, Störquellen und Details der Signalaufbereitung beeinflusst werden. Für geräte- oder zentrumsübergreifende Vergleiche erscheint es daher unerlässlich, diese Parameter nur innerhalb klar definierter technischer Rahmenbedingungen und in Kombination mit weiteren Kennwerten heranzuziehen.

Auf übergeordneter Ebene eröffnet das Framework die Perspektive, Qualitätsstandards in der EKT stärker datenbasiert zu definieren. Eine einheitliche, digital gestützte Beschreibung der Krampfqualität erleichtert interne Qualitätszirkel, die Abstimmung zwischen Behandlern und die Zusammenarbeit zwischen Zentren. Sie schafft zugleich die Voraussetzung für strukturierte Verlaufsdokumentationen beim Einsatz neuer Stimulationsparadigmen. Auf dieser Basis werden prospektive Register, ergebnisorientierte Auswertungen und klinische Studien zu Dosis-Wirkungs-Beziehungen, Response-Verläufen und Nebenwirkungen besser planbar. Die vorliegende Arbeit legt damit einen Baustein für eine EKT-Praxis, in der Qualitätsparameter nicht nur retrospektiv dokumentiert, sondern aktiv zur Steuerung und Evaluation der Behandlung genutzt werden.

Trotz dieser Möglichkeiten bleibt das klinische Urteil zentraler Bezugspunkt der Entscheidungsfindung. Ein zukünftiger Einsatz des Frameworks ist vor allem als klinisches Entscheidungsunterstützungssystem (engl. *Clinical Decision Support System*) denkbar: Digi-

tale Kennwerte machen Auffälligkeiten sichtbar, strukturieren Verläufe und können Vorschläge für Parameteranpassungen oder weiterführende Abklärungen vorbereiten, ersetzen aber weder die ganzheitliche Beurteilung des psychischen und somatischen Zustands noch die individuelle Abwägung von Nutzen und Risiken. Ein sinnvoller Anwendungsmodus besteht darin, die automatisiert gewonnenen Metriken als zusätzliche, standardisierte Informationsquelle in bestehende Entscheidungsprozesse zu integrieren. Etwa bei der Anpassung der Stimulationsenergie, bei der Bewertung grenzwertiger Anfälle, bei der Planung von Therapiepausen oder in multiprofessionellen Fallbesprechungen. In dieser Rolle können digitale Qualitätsparameter und ärztliche Erfahrung sich wechselseitig verstärken, ohne in Konkurrenz zueinander zu treten.

1.4.6 Ausblick

Die vorliegenden Ergebnisse eröffnen mehrere Perspektiven für weiterführende Forschung und methodische Entwicklung. Ein erster zentraler Schritt besteht in einer prospektiven multizentrischen Validierung des Frameworks. Dabei sollten unterschiedliche Gerätetypen, Ableitungsmontagen und Versorgungsumgebungen einbezogen werden, um zu prüfen, in welchem Ausmaß sich Modelle und Qualitätsparameter auf andere technische und organisatorische Rahmenbedingungen übertragen lassen. Wichtig ist eine strikt patientenbasierte Aufteilung der Daten sowie eine unabhängige Erhebung von Vergleichsmaßen durch geschulte Beurteiler. Auf dieser Grundlage kann die Generalisierbarkeit der Endpunktbestimmung ebenso wie die Stabilität der zentralen Qualitätsparameter im größeren Maßstab bewertet werden.

Ein weiterer Entwicklungsschwerpunkt liegt in der Erweiterung des Eingangsdatensatzes. Die hier verwendete zweikanalige EEG-Ableitung bildet nur einen Ausschnitt der kortikalen Aktivität während der EKT ab. Multikanalige EEG-Systeme ermöglichen eine differenziertere räumliche Betrachtung des iktalen Geschehens und eröffnen Merkmalsfamilien, etwa zur intrazerebralen Ausbreitung oder zu regional unterschiedlichen Erholungsmustern. Die Integration solcher Daten in das bestehende Framework erlaubt die Prüfung, ob komplexere räumliche Muster zusätzliche Aussagekraft für die Bewertung der Krampfqualität oder für die Vorhersage klinischer Verläufe bieten. Ergänzend können weitere

Biosignale wie Elektrokardiogramm, Blutdruck, Sauerstoffsättigung oder motorische Aktivität einbezogen werden, um multimodale Modelle der Therapieantwort zu entwickeln, die neurophysiologische und somatische Informationen gemeinsam abbilden.

Eine dritte Perspektive betrifft die Verknüpfung der Krampfqualitätsmetriken mit patientenbezogenen Outcomes. Bisher steht vor allem die Übereinstimmung mit geräteinternen und visuellen Referenzen im Vordergrund. Zukünftige Arbeiten sollten systematisch untersuchen, in welchem Ausmaß einzelne oder kombinierte Kennwerte der iktalen und postiktalen Aktivität mit klinischem Ansprechen, kognitiven Nebenwirkungen und langfristigen Verläufen assoziiert sind. Prospektive Studien mit standardisierten Outcome Erhebungen können dazu beitragen, Schwellenwerte und Zielbereiche für die Krampfqualität zu definieren und zu prüfen, ob sich bestimmte Muster als prädiktive Marker für Remission oder für ein erhöhtes Risiko kognitiver Beeinträchtigungen eignen. Auf dieser Basis lassen sich adaptivere Stimulationsstrategien und prädiktive Entscheidungsmodelle entwickeln, die über die reine Beschreibung des Krampfgeschehens hinausgehen und EKT an objektivierten Zielgrößen ausrichten.

Schließlich bietet die modulare Struktur des Frameworks Raum für methodische Weiterentwicklungen. Neue Merkmalsdefinitionen, alternative Endpunktkonzepte und weiterentwickelte Klassifikationsverfahren können in kontrollierter Weise in die bestehende Architektur integriert und mit den etablierten Ansätzen verglichen werden. Denkbar ist auch eine sukzessive Verschiebung von rein retrospektiven Analysen hin zu Anwendungen in Echtzeit, bei denen ausgewählte Kennwerte bereits im Behandlungsraum zur Verfügung stehen und unmittelbar in die klinische Entscheidungsfindung einfließen. Voraussetzung hierfür sind robuste Implementierungen, klar definierte Anforderungen an Datenqualität und eine enge Einbindung in bestehende klinische Arbeitsabläufe. Die vorliegende Arbeit liefert die konzeptionelle und technische Grundlage, auf der solche Schritte geplant und in Kooperation mit unterschiedlichen Zentren umgesetzt werden können. Damit wird ein Weg zu einer EKT-Praxis skizziert, in der Qualitätsmetriken und klinisches Urteil systematisch miteinander verbunden werden.

1.4.7 Schlussfolgerung

Die vorliegende Arbeit belegt, dass eine standardisierte digitale Auswertung zweikanaliger EKT EEG Daten eine zuverlässige Bestimmung des Anfallsendpunkts und eine klinisch anschlussfähige Berechnung zentraler Qualitätsparameter ermöglicht. Damit wird die bisher vorwiegend geräte- und beobachterabhängige Beurteilung des Krampfgeschehens um einen reproduzierbaren, datengetragenen Rahmen ergänzt, der als Grundlage für multizentrische Validierungsstudien, outcomeorientierte Forschung und eine schrittweise Integration eines klinischen Entscheidungsunterstützungssystems (engl. *Clinical Decision Support System*) in die Versorgungspraxis dienen kann.

1.5 Zusammenfassung

Diese Arbeit markiert einen Schritt von der überwiegend manuellen und ratersensitiven Anfallsbewertung hin zu einer standardisierten und datengetragenen Beurteilung der Krampfqualität im Rahmen der Elektrokonvulsionstherapie. Ziel war es zu prüfen, ob EKT induzierte Krampfanfälle auf Grundlage zweikanaliger EEG-Aufzeichnungen mit Methoden des maschinellen Lernens zuverlässig erkannt und etablierte Qualitätsparameter automatisiert berechnet werden können.

Ausgehend von 1.634 EKT-Behandlungen an 116 Patienten in zwei universitären Zentren wurde eine standardisierte Analysepipeline aufgebaut. Die aufgezeichneten EEG-Signale wurden segmentiert, mit einem überwachten Lernverfahren als „iktal“ oder „nicht-iktal“ klassifiziert und zur Bestimmung eines einheitlich definierten Anfallsendpunkts genutzt. Auf dieser Grundlage wurden Krampfdauer, postiktaler Suppressionsindex sowie energie- und amplitudenbasierte Kennwerte und kohärenzbasierte Maße berechnet und mit den Referenzangaben des Stimulationsgeräts und visuellen Expertenbewertungen verglichen.

Die Ergebnisse zeigen, dass die segmentbasierte Klassifikation eine stabile Trennung iktaler und nicht-iktaler Aktivität ermöglicht und einen konsistenten Endpunkt des Anfallsverlaufs liefert. Krampfdauer sowie energie- und amplitudenbezogene Indizes können durch das Framework in einer Qualität bestimmt werden, die eng an bestehende klinische

Bewertungsgewohnheiten anschließt. Einzelne Kennwerte wie der postiktale Suppressionsindex und kohärenzbasierte Maße erweisen sich dagegen als stärker abhängig von Definitionskonventionen und technischen Rahmenbedingungen und sind daher vorsichtiger zu interpretieren.

Damit beantwortet die Arbeit die in der Einleitung formulierten Fragestellungen. Sie zeigt, dass eine digital gestützte Anfallserkennung und Krampfqualitätsbewertung in der Elektrokonvulsionstherapie grundsätzlich zuverlässig umsetzbar ist und dass zentrale Qualitätsparameter automatisiert und reproduzierbar berechnet werden können. Die modulare Struktur des Frameworks bietet zudem einen Rahmen, in dem weitere Merkmalsfamilien und Datenquellen integriert sowie Zusammenhänge zwischen Krampfqualitätsprofilen und klinischen Outcomes systematisch untersucht werden können.

Insgesamt legt die Dissertation eine konzeptionelle und technische Grundlage für eine standardisierte digitale Bewertung der Krampfqualität in der Elektrokonvulsionstherapie. Auf dieser Basis sind prospektive multizentrische Validierungsstudien, multimodale Erweiterungen der Datengrundlage und die Entwicklung prädiktiver Entscheidungsinstrumente mögliche nächste Schritte, um die EKT als wichtiges Therapiemodul bei schwer behandelbaren psychischen Erkrankungen weiter zu optimieren.

1.6 Literaturverzeichnis der deutschen Zusammenfassung

Abbott CC, Loo D, Sui J. Determining Electroconvulsive Therapy Response With Machine Learning. JAMA Psychiatry 2016;73: 545

Accornero F. An Eyewitness Account of the Discovery of Electroshock. Convuls Ther 1988;4: 40–49

Al-Kaysi AM, Al-Ani A, Galvez V, Loo CK, Ling S, Boonstra TW. Estimating The Quality of Electroconvulsive Therapy Induced Seizures Using Decision Tree and Fuzzy Inference System Classifiers. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Honolulu, HI: IEEE, 2018: 3677–3680 https://ieeexplore.ieee.org/document/8513334/ (Zugriffsdatum: 4 Dezember 2023)

Aloysi AS, Bryson EO, Kellner CH. Management of Prolonged Seizures During Electroconvulsive Therapy. Indian J Psychol Med 2014;36: 220–221

Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Stat Surv 2010;4 https://projecteuclid.org/journals/statistics-surveys/volume-4/issue-none/A-survey-of-cross-validation-procedures-for-model-selection/10.1214/09-SS054.full (Zugriffsdatum: 9 März 2025)

Aten JJ, Oudega M, Van Exel E, Stek ML, Van Waarde JA. Repeated dose titration versus age-based method in electroconvulsive therapy: a pilot study. Eur Arch Psychiatry Clin Neurosci 2015;265: 351–356

Aytuluk HG, Simsek T, Yilmaz M, Turan AZ, Saracoglu KT. Can Propofol Lead to an Increase in Seizure Threshold Over the Course of Electroconvulsive Therapy? Clin Psychopharmacol Neurosci 2019;17: 523–530

Azuma H, Fujita A, Sato K, Arahata K, Otsuki K, Hori M, Mochida Y, Uchida M, Yamada T, Akechi T, Furukawa TA. Postictal suppression correlates with therapeutic efficacy for depression in bilateral sine and pulse wave electroconvulsive therapy. Psychiatry Clin Neurosci 2007;61: 168–173

Azuma H, Ogawa H, Suzuki E, Akechi T. Intraclass correlations of seizure duration by wavelet transform, sample entropy, and visual determination in electroconvulsive therapy. Neuropsychopharm Rep 2020;40: 102–106

Baghai TC, Frey R, Kasper S, Möller H-J, Herausgeber. Elektrokonvulsionstherapie: Klinische und wissenschaftliche Aspekte. Vienna: Springer Vienna, 2004 https://link.springer.com/10.1007/978-3-7091-3752-9 (Zugriffsdatum: 4 März 2025)

Bailine SH, Petrides G, Doft M, Lui G. Indications for the Use of Propofol in Electroconvulsive Therapy: J ECT 2003;19: 129–132

BÄK, KBV, AWMF. Nationale VersorgungsLeitlinie Unipolare Depression - Langfassung. Bundesärztekammer (BÄK); Kassenärztliche Bundesvereinigung (KBV); Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF), 2022 https://www.leitlinien.de/themen/depression/pdf/depression-vers3-2-lang.pdf (Zugriffsdatum: 30 November 2025)

Barandas M, Folgado D, Fernandes L, Santos S, Abreu M, Bota P, Liu H, Schultz T, Gamboa H. TSFEL: Time Series Feature Extraction Library. SoftwareX 2020;11: 100456

Beale MD, Kellner CH, Pritchett JT, Bernstein HJ, Burns CM, Knapp R. Stimulus dose-titration in ECT: a 2-year clinical experience. Convuls Ther 1994;10: 171–176

Beyer JL, Weiner RD, Glenn MD. Electroconvulsive therapy: a programmed text. 2nd ed. Washington, DC: American Psychiatric Press, 1998

Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit 1997;30: 1145–1159

Breiman L. Random Forests. Mach Learn 2001;45: 5–32

Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification And Regression Trees. 1. Aufl. Routledge, 2017 https://www.taylorfrancis.com/books/9781351460491 (Zugriffsdatum: 9 März 2025)

Busner J, Targum SD. The Clinical Global Impressions Scale. Psychiatry (Edgmont) 2007;4: 28–37

Bzdok D, Meyer-Lindenberg A. Machine Learning for Precision Psychiatry: Opportunities and Challenges. Biol Psychiatry Cogn Neurosci Neuroimaging 2018;3: 223–230

Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 2020;21: 6

Christensen P, Koldbæk I. EEG Monitored ECT. Br J Psychiatry 1982;141: 19–23

Cochran WG. THE COMPARISON OF PERCENTAGES IN MATCHED SAMPLES. Biometrika 1950;37: 256–266

Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20: 273–297

Cover T, Hart P. Nearest neighbor pattern classification. IEEE Trans Inf Theory 1967;13: 21–27

Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ. Random forests for classification in ecology. Ecology 2007;88: 2783–2792

De Arriba-Arnau A, Dalmau Llitjos A, Soria V, Savino S, Salvat-Pujol N, Curto J, Menchón JM, Urretavizcaya M. Factors Predicting Ictal Quality in Bilateral Electroconvulsive Therapy Sessions. Brain Sci 2021;11: 781

Delva NJ, Brunet D, Hawken ER, Kesteven RM, Lawson JS, Lywood DW, Rodenburg M, Waldron JJ. Electrical Dose and Seizure Threshold: Relations to Clinical Outcome and Cognitive Effects in Bifrontal, Bitemporal, and Right Unilateral ECT: J ECT 2000;16: 361–369

Dietterich T. Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. Neural Comput 1998;10: 1895–1923

Duda RO, Stork DG, Hart PE. Pattern classification. 2nd ed. New York: Wiley, 2001

Duhamel P, Vetterli M. Fast Fourier transforms: A tutorial review and a state of the art. Signal Processing Elsevier, 1990;19: 259–299

Edwards AL. Note on the “correction for continuity” in testing the significance of the difference between correlated proportions. Psychometrika 1948;13: 185–187

Espinoza RT, Kellner CH. Electroconvulsive Therapy. Ropper AH, Hrsg. N Engl J Med 2022;386: 667–672

Farzan F, Boutros NN, Blumberger DM, Daskalakis ZJ. What Does the Electroencephalogram Tell Us About the Mechanisms of Action of ECT in Major Depressive Disorders? J ECT 2014;30: 98–106

Fawcett T. An introduction to ROC analysis. Pattern Recognit Lett 2006;27: 861–874

Fink M. Relation of Electroencephalographic Delta Activity to Behavioral Response in Electroshock: Quantitative Serial Studies. AMA Arch Neurol Psychiatry 1957;78: 516

Fink M. Meduna and the origins of convulsive therapy. Am J Psychiatry US: American Psychiatric Assn, 1984;141: 1034–1041

Fink M. Convulsive therapy: a review of the first 55 years. J Affect Disord 2001;63: 1–15

Francis-Taylor R, Ophel G, Martin D, Loo C. The ictal EEG in ECT: A systematic review of the relationships between ictal features, ECT technique, seizure threshold and outcomes. Brain Stimul 2020;13: 1644–1654

Freeman CP, Herausgeber. The ect handbook: the second report of the Royal College of Psychiatrists' Special Committee on ECT. London: Royal College of Psychiatrists, 1995

Freundlieb N, Schneider E, Brühl A, Kiebs M. GENET-GPD: A documentation tool to digitally collect longitudinal ECT treatment data and associated biosignals. Brain Stimulation 2023;16: 1173–1175

Friedman JH. Greedy function approximation: A gradient boosting machine. Ann. Statist. 2001;29 https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-5/Greedy-function-approximation-A-gradient-boosting-machine/10.1214/aos/1013203451.full (Zugriffsdatum: 9 März 2025)

Gaebel W, Hasan A, Falkai P. S3-Leitlinie Schizophrenie. Berlin, Heidelberg: Springer Berlin / Heidelberg, 2019

Gangadhar BN, Dutt DN, Janakiramaiah N, Sadasivan PK. Automation of seizure duration estimation during ECT; use of fractal dimension. Proceedings of the First Regional Conference, IEEE Engineering in Medicine and Biology Society and 14th Conference of the Biomedical Engineering Society of India. An International Meet. New Delhi, India: IEEE, 1995: 3/37-3/38 http://ieeexplore.ieee.org/document/532965/ (Zugriffsdatum: 14 März 2025)

Gärtner M, Ghisu E, Herrera-Melendez AL, Koslowski M, Aust S, Asbach P, Otte C, Regen F, Heuser I, Borgwardt K, Grimm S, Bajbouj M. Using routine MRI data of depressed patients to predict individual responses to electroconvulsive therapy. Exp Neurol 2021;335: 113505

Geisser S. The Predictive Sample Reuse Method with Applications. J Am Stat Assoc 1975;70: 320–328

Gillving C, Ekman CJ, Hammar Å, Landén M, Lundberg J, Movahed Rad P, Nordanskog P, Von Knorring L, Nordenskjöld A. Seizure Duration and Electroconvulsive Therapy in Major Depressive Disorder. JAMA Netw Open 2024;7: e2422738

Grözinger M, Conca A, Nickl-Jockschat T, Di Pauli J, Herausgeber. Elektrokonvulsionstherapie kompakt. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013 http://link.springer.com/10.1007/978-3-642-25629-5 (Zugriffsdatum: 7 Oktober 2022)

Grundmann U, Oest M. Anästhesiologische Aspekte bei Elektrokrampftherapie. Anaesthesist 2007;56: 202–211

Guze BH, Liston EH, Baxter LRJr, Richeimer SH, Gold ME. Poor interrater reliability of MECTA EEG recordings of ECT seizure duration. J Clin Psychiatry 1989;50: 140–142

Haesebaert J, Moreno A, Lesage A, Patry S, Lemasson M. A Descriptive Study of Data Collection Systems Used in Electroconvulsive Therapy Units in the Province of Quebec, Canada. J ECT 2020;36: 36–41

Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE. Array programming with NumPy. Nature Nature Publishing Group, 2020;585: 357–362

Hastie T, Friedman J, Tibshirani R. The Elements of Statistical Learning. New York, NY: Springer New York, 2001 http://link.springer.com/10.1007/978-0-387-21606-5 (Zugriffsdatum: 9 März 2025)

Hoppe J, Scriba P. Stellungnahme zur Elektrokrampftherapie (EKT) als psychiatrische Maßnahme. Dtsch Arztebl 2003;100: A142–A143

Itani S, Rossignol M. At the Crossroads Between Psychiatry and Machine Learning: Insights Into Paradigms and Challenges for Clinical Applicability. Front Psychiatry 2020;11: 552262

Janouschek H, Langbehn DR, Nickl-Jockschat T, Grözinger M. The impact of seizure quality on ect treatment efficacy. Psychiatry Res 2020;293: 113466

Joblib Development Team. Joblib: running Python functions as pipeline jobs. 2020 https://joblib.readthedocs.io/

Kayser M. ECTMetrics: Python Library for Calculating Seizure Quality Indices in Electroconvulsive Therapy. Zenodo, 2025 https://zenodo.org/doi/10.5281/zenodo.14285753 (Zugriffsdatum: 29 Januar 2025)

Kayser M, Hurlemann R, Philipsen A, Freundlieb N, Kiebs M. Machine learning based seizure classification and digital biosignal analysis of ECT seizures. Sci Rep 2025;15: 6409

Kellner CH, Li D, Maron L. Anesthesia for Electroconvulsive Therapy. In: Swartz CM, Hrsg. Electroconvulsive and Neuromodulation Therapies. New York: Cambridge University Press, 2009: 412–429

Kemp B, Olivan J. European data format 'plus' (EDF+), an EDF alike standard format for the exchange of physiological data. Clin Neurophysiol 2003;114: 1755–1761

Kemp B, Värri A, Rosa AC, Nielsen KD, Gade J. A simple format for exchange of digitized polygraphic recordings. Electroencephalogr Clin Neurophysiol 1992;82: 391–393

Kiebs M, Freundlieb N. GENET - A novel international research collaboration for the standardized collection and sharing of ECT treatment data and associated biosignals – Implementation, Application, and Findings. Brain Stimul 2021;14: 1737–1738

Kramer BA, Pollock VE, Schneider LS, Gray GE. Interrater reliability of MECTA SR-1 seizure duration. Biol Psychiatry 1989;25: 642–644

Kranaster L, Janke C, Lewien A, Hoyer C, Lederbogen F, Sartorius A. Rethinking Restimulation: A Case Report. J ECT 2012;28: 248–249

Krystal AD, Coffey CE, Weiner RD, Holsinger T. Changes in Seizure Threshold Over the Course of Electroconvulsive Therapy Affect Therapeutic Response and Are Detected by Ictal EEG Ratings. J Neuropsychiatry Clin Neurosci 1998;10: 178–186

Krystal AD, Weiner RD. ECT Seizure Duration: Reliability of Manual and Computer-Automated Determinations. Convuls Ther 1995;11: 158–169

Krystal AD, Weiner RD, Coffey CE. The ictal EEG as a marker of adequate stimulus intensity with unilateral ECT. J Neuropsychiatry Clin Neurosci 1995;7: 295–303

Krystal AD, Weiner RD, Lindahl V, Massie R. The Development and Retrospective Testing of an Electroencephalographic Seizure Quality–Based Stimulus Dosing Paradigm With ECT. J ECT 2000;16 https://journals.lww.com/ectjournal/fulltext/2000/12000/the_development_and_retrospective_testing_of_an.3.aspx

Krystal AD, Weiner RD, McCall WV, Shelp FE, Arias R, Smith P. The effects of ECT stimulus dose and electrode placement on the Ictal electroencephalogram: An intraindividual crossover study. Biol Psychiatry 1993;34: 759–767

Kubat M. An Introduction to Machine Learning. Cham: Springer International Publishing, 2015

Lemaitre G, Nogueira F, Aridas CK. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. arXiv, 2016 https://arxiv.org/abs/1609.06570 (Zugriffsdatum: 30 November 2025)

Levene H. Robust tests for equality of variances. Contributions to probability and statistics : essays in honor of Harold Hotelling. Stanford, California: Stanford University Press, 1960

Liaw A, Wiener M. Classification and Regression by randomForest. R News 2002;2: 18–22

Liljequist D, Elfving B, Skavberg Roaldsen K. Intraclass correlation – A discussion and demonstration of basic features. Chiacchio F, Hrsg. PLoS One 2019;14: e0219854

Lipton ZC, Elkan C, Naryanaswamy B. Optimal Thresholding of Classifiers to Maximize F1 Measure. In: Calders T, Esposito F, Hüllermeier E, Meo R, Hrsg. Machine Learning and Knowledge Discovery in Databases. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014: 225–239 http://link.springer.com/10.1007/978-3-662-44851-9_15 (Zugriffsdatum: 10 März 2025)

Little RJA, Rubin DB. Statistical analysis with missing data. Third edition. Hoboken, NJ: Wiley, 2020

Loh W-Y. Classification and regression trees. WIREs Data Min Knowl Discov 2011;1: 14–23

Lundin RM, Falcao VP, Kannangara S, Eakin CW, Abdar M, O'Neill J, Khosravi A, Eyre H, Nahavandi S, Loo C, Berk M. Machine Learning in Electroconvulsive Therapy: A Systematic Review. J ECT 2024 https://journals.lww.com/10.1097/YCT.0000000000001009 (Zugriffsdatum: 13 August 2024)

Lundin RM, Menkes DB. Realising the potential of digital psychiatry. Lancet Psychiatry 2021;8: 655

Maharana K, Mondal S, Nemade B. A review: Data pre-processing and data augmentation techniques. Global Transitions Proceedings 2022;3: 91–99

Mann HB, Whitney DR. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. The Annals of Mathematical Statistics 1947;18: 50–60

Mason L, Baxter J, Bartlett P, Frean M. Boosting Algorithms as Gradient Descent. 1999: 512–518

Matthews BW. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim Biophys Acta 1975;405: 442–451

Mayur P. Ictal Electroencephalographic Characteristics During Electroconvulsive Therapy: A Review of Determination and Clinical Relevance. J ECT 2006;22: 213–217

McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Med 1996;1: 30–46

McKinney W. Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference 2010: 56–61

McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 1947;12: 153–157

Mohammed R, Rawashdeh J, Abdullah M. Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results. 2020 11th International Conference on Information and Communication Systems (ICICS). Irbid, Jordan: IEEE, 2020: 243–248 https://ieeexplore.ieee.org/document/9078901/ (Zugriffsdatum: 15 März 2025)

Morrens J, Mathews M, Popova V, Borentain S, Rive B, Gonzalez Martin Moro B, Jamieson C, Zhang Q. Use of Clinical Global Impressions-Severity (CGI-S) to Assess Response to Antidepressant Treatment in Patients with Treatment-Resistant Depression. Neuropsychiatr Dis Treat 2022;Volume 18: 1127–1132

Nahrstaedt H, Kern S, LucaCerina, Appelhoff S, Kao DTH, Franklin C, Clarke S, Zitting J, Ojeda D, Boulay C, Ogi Moore, Kosachev D. holgern/pyedflib: v0.1.23. Zenodo, 2020 https://zenodo.org/record/5678481 (Zugriffsdatum: 7 Oktober 2022)

Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobot 2013;7 http://journal.frontiersin.org/article/10.3389/fnbot.2013.00021/abstract (Zugriffsdatum: 9 März 2025)

Niemantsverdriet L, Birkenhäger TK, Van Den Broek WW. The Efficacy of Ultrabrief-Pulse (0.25 millisecond) Versus Brief-Pulse (0.50 millisecond) Bilateral Electroconvulsive Therapy in Major Depression. J ECT 2011;27: 55–58

Nobler MS, Luber B, Moeller JR, Katzman GP, Prudic J, Devanand DP, Dichter GS, Sackeim HA. Quantitative EEG During Seizures Induced by Electroconvulsive Therapy: Relations to Treatment Modality and Clinical Features. I. Global Analyses: J ECT 2000;16: 211–228

Nobler MS, Sackeim HA, Solomou M, Luber B, Devanand DP, Prudic J. EEG manifestations during ECT: effects of electrode placement and stimulus intensity. Biol Psychiatry 1993;34: 321–330

Olbrich S, Arns M. EEG biomarkers in major depressive disorder: Discriminative power and prediction of treatment response. Int Rev Psychiatry 2013;25: 604–618

Omohundro SM. Five Balltree Construction Algorithms. International Computer Science Institute, 1989Report No.: TR-89-063

Pagnin D, de Queiroz V, Pini S, Cassano GB. Efficacy of ECT in depression: a meta-analytic review. J ECT 2004;20: 13–20

Pape A, Kittel-Schneider S. Praxis der Anästhesie für die Elektrokrampftherapie. Anästhesiol Intensivmed Notfallmed Schmerzther 2023;58: 422–435

Pearson K. LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science Taylor & Francis, 1901;2: 559–572

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. Scikit-learn: Machine Learning in Python. J Mach Learn Res 2011;12: 2825–2830

Perera TD, Coplan JD, Lisanby SH, Lipira CM, Arif M, Carpio C, Spitzer G, Santarelli L, Scharf B, Hen R, Rosoklija G, Sackeim HA, Dwork AJ. Antidepressant-Induced Neurogenesis in the Hippocampus of Adult Nonhuman Primates. J Neurosci 2007;27: 4894–4901

Perera TD, Luber B, Nobler MS, Prudic J, Anderson C, Sackeim HA. Seizure Expression During Electroconvulsive Therapy: Relationships with Clinical Outcome and Cognitive Side Effects. Neuropsychopharmacol 2004;29: 813–825

Powers DMW. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv, 2020 https://arxiv.org/abs/2010.16061 (Zugriffsdatum: 10 März 2025)

Prudic J, Olfson M, Sackeim HA. Electro-convulsive therapy practices in the community. Psychol Med 2001;31: 929–934

Quinlan JR. Induction of Decision Trees. Mach Learn 1986;1: 81–106

Raschka S. MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. J Open Source Softw 2018;3: 638

Rasmussen K. The Practice of Electroconvulsive Therapy: Recommendations for Treatment, Training, and Privileging (Second Edition).: J ECT 2002;18: 58–59

Reback J, Jbrockmendel, McKinney W, Van Den Bossche J, Augspurger T, Cloud P, Hawkins S, Roeschke M, Gfyoung, Sinhrks, Klein A, Terji Petersen, Hoefler P, Tratner J, She C, Ayd W, Naveh S, Garcia M, JHM Darbyshire, Schendel J, Hayden A, Shadrach R, Saxton D, Gorelli ME, Fangchen Li, Zeitlin M, Jancauskas V, McMaster A, Battiston P, Skipper Seabold. pandas-dev/pandas: Pandas 1.3.5. Zenodo, 2021 https://zenodo.org/record/5774815 (Zugriffsdatum: 26 Dezember 2023)

Redlich R, Opel N, Grotegerd D, Dohm K, Zaremba D, Bürger C, Münker S, Mühlmann L, Wahl P, Heindel W, Arolt V, Alferink J, Zwanzger P, Zavorotnyy M, Kugel H, Dannlowski U. Prediction of Individual Response to Electroconvulsive Therapy via Machine Learning on Structural Magnetic Resonance Imaging Data. JAMA Psychiatry 2016;73: 557

Ries RK. Poor interrater reliability of MECTA EEG seizure duration measurement during ECT. Biol Psychiatry 1985;20: 94–98

Roemer RA, Shagass C, Dubin W, Jaffe R, Katz R. Relationship between Pretreatment Electroencephalographic Coherence Measures and Subsequent Response to Electroconvulsive Therapy: A Preliminary Study. Neuropsychobiology 1990;24: 121–124

Rokhlin V, Szlam A, Tygert M. A randomized algorithm for principal component analysis. arXiv, 2008 https://arxiv.org/abs/0809.2274 (Zugriffsdatum: 15 März 2025)

Rosenquist PB, McCall WV, Colenda CC, Melton BA. A Comparison of Visual and Computer-Generated Measures of “Seizure Quality”. J ECT 1998;14: 76–82

Rossum G van, Drake FL. The Python language reference. Release 3.0.1 [Repr.]. Hampton, NH: Python Software Foundation, 2010

Sackeim HA. The Effects of Electroconvulsive Therapy on Quantitative Electroencephalograms: Relationship to Clinical Outcome. Arch Gen Psychiatry 1996;53: 814

Sackeim HA, Devanand DP, Prudic J. Stimulus Intensity, Seizure Threshold, and Seizure Duration. Psychiatr Clin North Am 1991;14: 803–843

Sackeim HA, Prudic J, Devanand DP, Nobler MS, Lisanby SH, Peyser S, Fitzsimons L, Moody BJ, Clark J. A Prospective, Randomized, Double-blind Comparison of Bilateral and Right Unilateral Electroconvulsive Therapy at Different Stimulus Intensities. Arch Gen Psychiatry 2000;57: 425

Sanacora G, Mason GF, Rothman DL, Hyder F, Ciarcia JJ, Ostroff RB, Berman RM, Krystal JH. Increased Cortical GABA Concentrations in Depressed Patients Receiving ECT. Am J Psychiatry 2003;160: 577–579

Saripuddin M, Suliman A, Syarmila Sameon S, Jorgensen BN. Random Undersampling on Imbalance Time Series Data for Anomaly Detection. 2021 The 4th International Conference on Machine Learning and Machine Intelligence. Hangzhou China: ACM, 2021: 151–156 https://dl.acm.org/doi/10.1145/3490725.3490748 (Zugriffsdatum: 15 März 2025)

Sasaki Y. The truth of the F-measure. Teach Tutor Mater 2007

Scangos KW, Weiner RD, Coffey EC, Krystal AD. An Electrophysiological Biomarker That May Predict Treatment Response to ECT. J ECT 2019;35: 95–102

Schneider F, Härter M, Schorr S, Herausgeber. S3-Leitlinie/Nationale VersorgungsLeitlinie Unipolare Depression. 2. Aufl. 2017. Berlin, Heidelberg: Springer Berlin Heidelberg, 2017

Scott AIF. Monitoring electroconvulsive therapy by electroencephalogram: an update for ECT practitioners. Adv Psychiatr Treat 2007;13: 298–304

Shapiro SS, Wilk MB. An Analysis of Variance Test for Normality (Complete Samples). Biometrika 1965;52: 591

Shaw JC. An introduction to the coherence function and its use in EEG signal analysis. J Med Eng Technol 1981;5: 279–288

Sheskin DJ. Handbook of parametric and nonparametric statistical procedures. Fifth edition. Boca Raton London New York: CRC Press, Taylor & Francis Group, 2011

Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 1979;86: 420–428

Simon L, Blay M, Galvao F, Brunelin J. Using EEG to Predict Clinical Response to Electroconvulsive Therapy in Patients With Major Depression: A Comprehensive Review. Front Psychiatry 2021;12: 643710

Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Inf Process Manag 2009;45: 427–437

Somatics LLC. Thymatron® System IV Instructions for Use. 720 Commerce Drive, Suite 101, Venice, FL 34292, United States of America, 2019 https://www.thymatron.com/downloads/System_IV_Instruction_Manual_Rev_21.pdf

Sørensen P, Bolwig TG, Lauritsen B, Bengtson O. Electroconvulsive therapy: a comparison of seizure duration as monitored with electroencephalograph and electromyograph. Acta Psychiatr Scand 1981;64: 193–198

Spearman C. The Proof and Measurement of Association between Two Things. Am J Psychol 1904;15: 72

Squire LR, Zouzounis JA. ECT and memory: brief pulse versus sine wave. Am J Psychiatry 1986;143: 596–601

Sun H, Jiang R, Qi S, Narr KL, Wade BS, Upston J, Espinoza R, Jones T, Calhoun VD, Abbott CC, Sui J. Preliminary prediction of individual response to electroconvulsive therapy using whole-brain functional magnetic resonance imaging data. Neuroimage Clin 2020;26: 102080

Suppes T, Webb A, Carmody T, Gordon E, Gutierrez-Esteinou R, Hudson JI, Pope HG. Is postictal electrical silence a predictor of response to electroconvulsive therapy? J Affect Disord 1996;41: 55–58

Swartz CM. Physiological response to ECT stimulus dose. Psychiatry Res 2000;97: 229–235

Swartz CM. Stimulus Electrode Placement. In: Swartz CM, Hrsg. Electroconvulsive and Neuromodulation Therapies. New York: Cambridge University Press, 2009: 430–446

Swartz CM, Abrams R, Rasmussen K, Pavel J, Zorumski CF, Srinivasaraghavan J. Computer Automated Versus Visually Determined Electroencephalographic and Electromyographic Seizure Duration. Convuls Ther 1994;10: 165–170

Swartz CM, Michael N. Age-Based Seizure Threshold Determination. J ECT 2013;29: 18–20

Ten Doesschate F, Van Wingen GA, De Pont BJHB, Arns M, Van Waarde JA. The Longitudinal Effects of Electroconvulsive Therapy on Ictal Interhemispheric Coherence and Its Associations With Treatment Outcome: A Naturalistic Cohort Study. Clin EEG Neurosci 2019;50: 44–50

Triantafyllidis AK, Tsanas A. Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature. J Med Internet Res 2019;21: e12286

Vallat R. Pingouin: statistics in Python. J Open Source Softw 2018;3: 1026

Vapnik VN. Statistical learning theory. New York: Wiley, 1998

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods Nature Publishing Group, 2020;17: 261–272

de Vreede IM, Burger H, van Vliet IM. Prediction of response to ECT with routinely collected data in major depression. J Affect Disord 2005;86: 323–327

van Waarde JA, van Oudheusden LJB, Verwey B, Giltay EJ, van der Mast RC. Clinical predictors of seizure threshold in electroconvulsive therapy: a prospective study. Eur Arch Psychiatry Clin Neurosci 2013;263: 167–175

Warmflash VL, Stricks L, Sackeim HA, Decina P, Neeley P, Malitz S. Reliability and Validity of Measures of Seizure Duration. Convuls Ther 1987;3: 18–25

Weiner RD, Coffey CE, Krystal AD. The Monitoring and Management of Electrically Induced Seizures. Psychiatr Clin North Am 1991;14: 845–869

Weiner RD, Reti IM. Key updates in the clinical application of electroconvulsive therapy. Int Rev Psychiatry 2017;29: 54–62

Wilimitis D, Walsh CG. Practical Considerations and Applied Examples of Cross-Validation for Model Development and Evaluation in Health Care: Tutorial. JMIR AI 2023;2: e49023

Yalcin S, Aydoğan H, Selek S, Kucuk A, Yuce HH, Karababa F, Bilgiç T. Ketofol in electroconvulsive therapy anesthesia: two stones for one bird. J Anesth 2012;26: 562–567

Yoo CS, Jung DC, Ahn YM, Kim YS, Kim S-G, Yoon H, Lim YJ, Yi SH. Automatic detection of seizure termination during electroconvulsive therapy using sample entropy of the electroencephalogram. Psychiatry Res 2012;195: 76–82

Zandi PP, Morreale M, Reti IM, Maixner DF, McDonald WM, Patel PD, Achtyes E, Bhati MT, Carr BR, Conroy SK, Cristancho M, Dubin MJ, Francis A, Glazer K, Ingram W, Khurshid K, McClintock SM, Pinjari OF, Reeves K, Rodriguez NF, Sampson S, Seiner SJ, Selek S, Sheline Y, Smetana RW, Soda T, Trapp NT, Wright JH, Husain M, Weiner RD. National Network of Depression Centers' Recommendations on Harmonizing Clinical Documentation of Electroconvulsive Therapy. J ECT 2022;38: 159–164

Zou KH, O'Malley AJ, Mauri L. Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models. Circulation 2007;115: 654–657

2. Veröffentlichung

Dieser Publikationsdissertation liegt die folgende, unabhängig begutachtete Veröffentlichung zugrunde:

Kayser M, Hurlemann R, Philipsen A, Freundlieb N, Kiebs M. Machine learning based seizure classification and digital biosignal analysis of ECT seizures. Sci Rep. 2025; 15: 6409.

https://doi.org/10.1038/s41598-025-88238-3

3. Erklärung zum Eigenanteil

Die zugrunde liegende Arbeit wurde an der Klinik für Psychiatrie und Psychotherapie des Universitätsklinikums Bonn unter der Betreuung von Univ.-Prof. Dr. med. Alexandra Philipsen durchgeführt und im Rahmen des SCIMED-Promotionsstipendiums der Universität Bonn durch ein begleitendes Mentoring unterstützt. Die Dissertation basiert auf einer Publikation, bei der ich als alleiniger Erstautor geführt wurde und den ersten Entwurf des Manuskripts eigenständig verfasst habe.

Die Planung der wissenschaftlichen Arbeit umfasste die Konzeption der Studie, die Definition der Forschungsziele sowie die Auswahl und Festlegung der Machine-Learning-Algorithmen und der statistischen Auswertungsverfahren. Diese konzeptionelle Ausarbeitung und die Entwicklung des methodischen Ansatzes einschließlich der Hypothesen erfolgten federführend durch mich in enger wissenschaftlicher Zusammenarbeit mit Dr. Maximilian Kiebs und unter der übergeordneten Betreuung von Univ.-Prof. Dr. med. Alexandra Philipsen. In die Verfeinerung des Studienkonzepts und die strategische Ausrichtung der Forschungsfragen war zudem maßgeblich PD Dr. Nils Freundlieb eingebunden.

Die Datenerhebung beruhte auf klinischen Routinedaten und EEG-Biosignalen aus der Elektrokonvulsionstherapie, die an zwei Behandlungszentren erhoben wurden. Die Konzeption und übergeordnete Koordination der Datenakquise lagen bei Dr. Maximilian Kiebs. Ich war verantwortlich für die Planung und Durchführung der für die Studie erforderlichen klinischen Messungen, die technische Integration der Daten in die Analyseumgebung sowie die Sicherstellung und Validierung der Datenqualität in enger Zusammenarbeit mit den klinischen Partnern. Bei der praktischen Datenerhebung und Dokumentation wurde ich durch wissenschaftliches und klinisches Personal unterstützt.

Die Auswertung der Daten, insbesondere die Implementierung, Optimierung und Validierung der Machine-Learning-Modelle sowie die Durchführung sämtlicher statistischer Analysen, erfolgte überwiegend eigenständig durch mich in enger fachlicher Abstimmung mit Dr. Maximilian Kiebs. Dies schließt die Vorbereitung der Datensätze, die Merkmalsextraktion aus den EEG-Signalen, die Modelltrainings- und Validierungsschritte sowie die Berechnung und den Vergleich der Krampfqualitätsparameter ein. Die Erstellung der in der

Publikation und in dieser Dissertation dargestellten Tabellen und Abbildungen zur Veranschaulichung der Ergebnisse wurde ebenfalls von mir vorgenommen.

Im Rahmen der zugrunde liegenden Originalpublikation lagen die Interpretation der Ergebnisse im Kontext des aktuellen Forschungsstandes sowie die Ableitung der klinischen Relevanz und der sich hieraus ergebenden Perspektiven für zukünftige Forschungsvorhaben in meiner Hauptverantwortung. Meine Promotionsbetreuerin und Koautorin, Univ.-Prof. Dr. med. Alexandra Philipsen, sowie die weiteren Koautoren Dr. Maximilian Kiebs und PD Dr. Nils Freundlieb haben durch kritisches Feedback, die Diskussion der Ergebnisse und inhaltliche Anmerkungen zur Präzisierung der Argumentation beigetragen. Darüber hinaus gingen Rückmeldungen externer Gutachterinnen und Gutachter im Rahmen des Peer-Review-Prozesses sowie Kommentare von Fachkolleginnen und Fachkollegen auf wissenschaftlichen Tagungen in die Überarbeitung des Manuskripts ein.

Im Zuge der Arbeit habe ich außerdem eigenständig eine Python-basierte Toolbox zur Berechnung von Qualitätsparametern entwickelt und als Open-Source-Repository für weitere Forschungszwecke bereitgestellt.

Ich versichere, die Dissertationsschrift selbständig verfasst zu haben und keine weiteren als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

4. Danksagung

Diese Dissertation wurde durch die Unterstützung und Begleitung vieler Personen und Institutionen ermöglicht, denen ich an dieser Stelle meinen tiefen Dank aussprechen möchte.

Zunächst danke ich meiner Doktormutter, Frau Univ.-Prof. Dr. med. Alexandra Philipsen, für den wissenschaftlichen Rahmen, ihr Vertrauen, ihre Expertise und ihre kontinuierliche Unterstützung während der Entstehung dieser Arbeit.

Ebenso danke ich Herrn Prof. Dr. Stephan Jonas für die Erstellung des Zweitgutachtens sowie Frau Priv.-Doz. Dr. med. Ana Kowark für die Durchführung der Wahlfachprüfung im Rahmen meiner mündlichen Promotionsprüfung.

Meinem Mentor und Co-Autor, Herrn Dr. Maximilian Kiebs, spreche ich meinen tief empfundenen Dank für seine unermüdliche Unterstützung aus. Seine fachliche Kompetenz, sein außergewöhnliches Engagement und seine verlässliche wissenschaftliche wie organisatorische Begleitung haben entscheidend zum Gelingen dieser Arbeit beigetragen.

Ebenso danke ich allen weiteren Co-Autoren, insbesondere Herrn PD Dr. Nils Freundlieb, für die konstruktive und stets inspirierende Zusammenarbeit, die Weiterentwicklung des Studienkonzepts und viele wertvolle Impulse bei der Interpretation der Ergebnisse.

Ein herzlicher Dank gilt allen Studienteilnehmerinnen und Studienteilnehmern, ohne deren Teilnahme diese Arbeit in dieser Form nicht möglich gewesen wäre.

Für die finanzielle Förderung durch das BONFOR Scimed-Promotionsstipendium sowie die Bereitstellung digitaler Ressourcen über das Bonna HPC-Cluster danke ich der Universität Bonn.

Abschließend gilt mein tiefster Dank meiner Familie und meinen Freunden, die mich während dieser intensiven Zeit mit Geduld und Zuspruch begleitet haben, sowie Felix, dessen Rückhalt mir in entscheidenden Momenten die Kraft gegeben hat, diese Arbeit zu vollenden.