Multivariate Verfahren

MSc Klinische Psychologie und Psychotherapie

MSc Umweltpsychologie/Mensch-Technik-Interaktion

MSc Psychologie

Prof. Dr. Dirk Ostwald | Wintersemester 2025/2026

Vorlesung Multivariate Verfahren WiSe 2025/2026

Datum	Abschnitt	Einheit	Lehrperson
17.10.2025	Grundlagen	(1) Einführung	DO, JS
20.10.2025	Grundlagen	(2) Matrizen	JS
31.10.2025	Reformationstag		
07.11.2025	Grundlagen	(3) Eigenanalyse	JS
14.11.2025	Grundlagen	(4) Multivariate Normalverteilungen	JS
21.11.2025	Frequentistische Inferenz	(5) Multivariate Deskriptivstatistik	JS
28.11.2025	Frequentistische Inferenz	(6) Multivariate Varianzanalyse	JS
05.12.2025	Frequentistische Inferenz	(7) Kanonische Korrelation	DO
12.12.2025	Prädiktive Modellierung	(8) Prädiktive Modellierung	DO
19.12.2025	Prädiktive Modellierung	(9) Dimensionsreduktion	DO
	Weihnachtspause		
09.01.2026	Prädiktive Modellierung	(10) Bayes-Klassifikation	JS
16.01.2026	Prädiktive Modellierung	(11) Nichtlineare Optimierung	DO
23.01.2026	Prädiktive Modellierung	(12) Logistische Regression	DO
30.01.2026	Prädiktive Modellierung	(13) Neuronale Netze	DO
Februar 2026	Klausurtermin	MSc KliPP	
Juli 2026	Klausurtermin	MSc UPsy/MTI und MSc Psychologie	
Juli 2026	Klausurwiederholungstermin	MSc KliPP	

(8) Prädiktive Modellierung

Rhetorik

Anwendungsbeispiel

Performanzmetriken

Selbstkontrollfragen

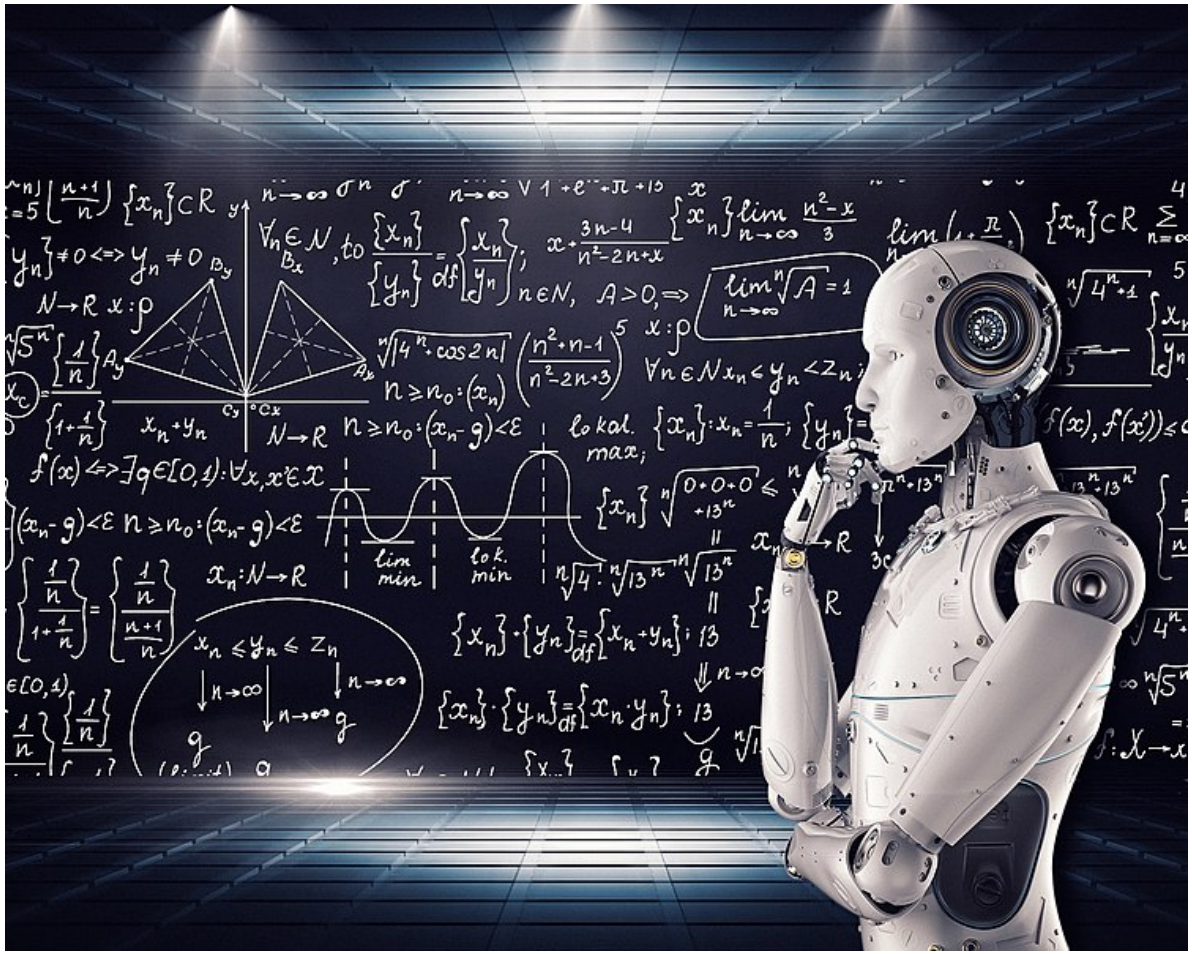
Rhetorik

Anwendungsbeispiel

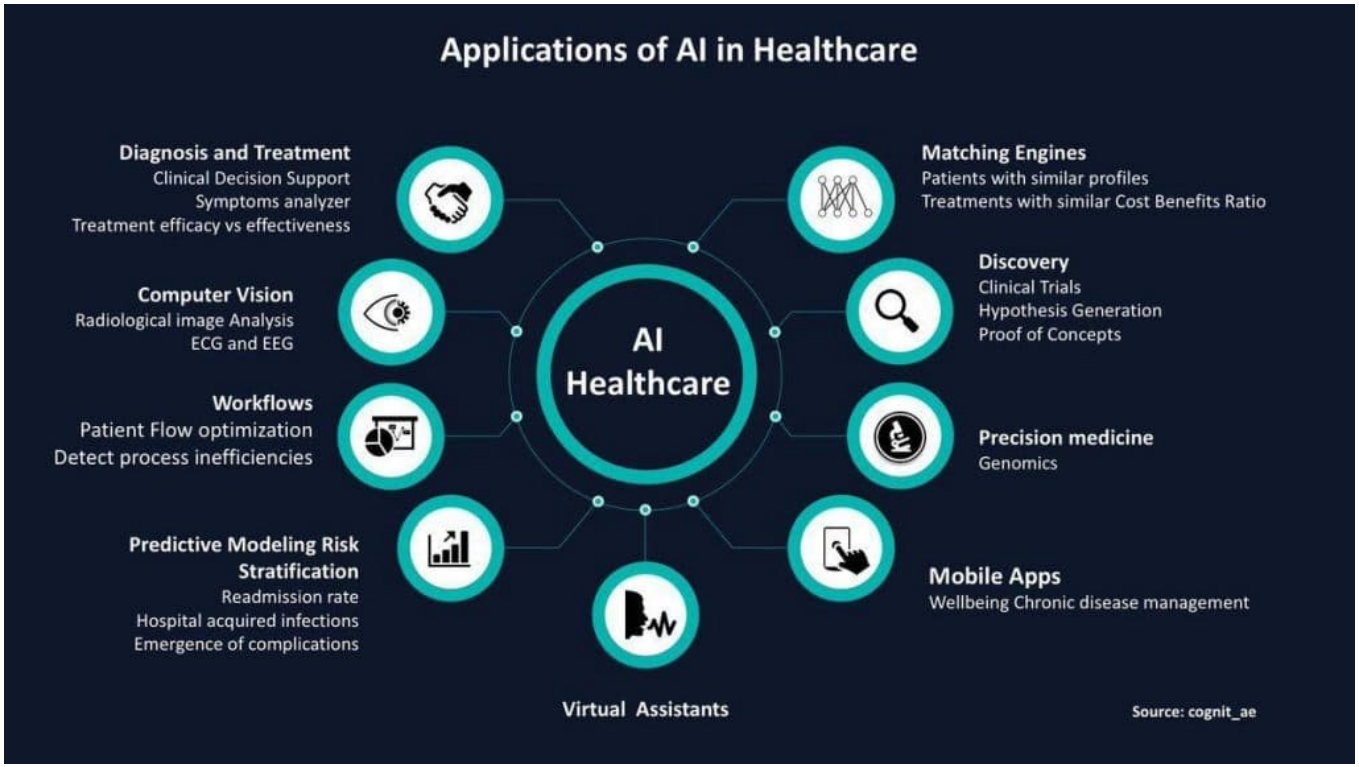
Performanzmetriken

Selbstkontrollfragen

Rhetorik

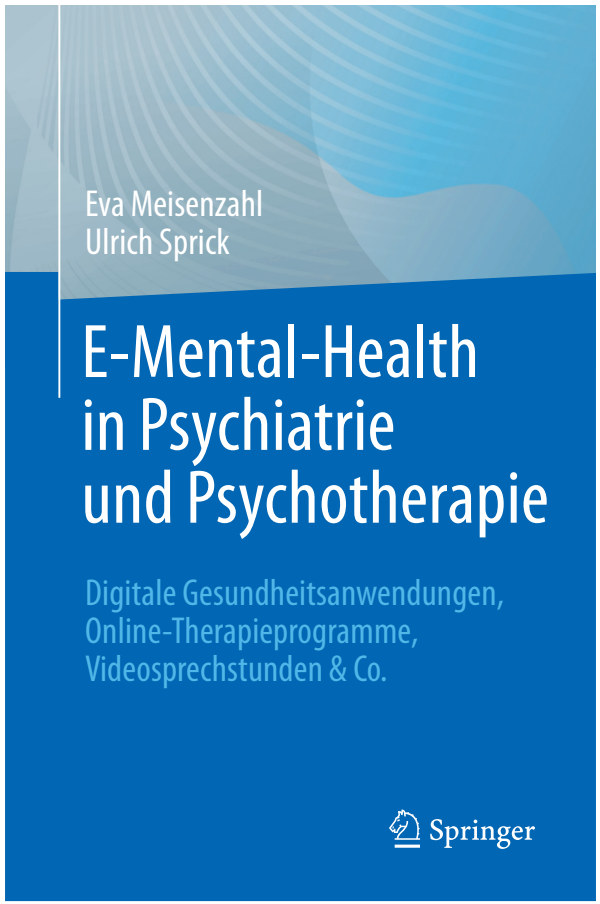


Wikimedia Commons



www.botreetechnologies.com

Pfannstiel (2022)



Meisenzahl & Sprick (2023)

DGPs | Deutsche Gesellschaft für Psychologie

Mitglieder-Login EN

Schwerpunkte Die DGPs Fachgruppen Psychologie studieren Presse Angebote Mitgliedschaft

AKTUELLES

Startseite > Aktuelles > DGPs-Kongress 2024: Wie Künstliche Intelligenz die Arbeitswelt, das Bildungswesen und die Psychotherapie verändert

DGPs-Kongress 2024: Wie Künstliche Intelligenz die Arbeitswelt, das Bildungswesen und die Psychotherapie verändert

11.09.2024 | Vorstand Pressemitteilung

Wie kann Künstliche Intelligenz (KI) gewinnbringend in der Schule eingesetzt werden? Welche Chancen und Risiken gibt es beim Einsatz von KI in der Arbeitswelt? Wie wird KI psychotherapeutische Diagnostik und Interventionen verändern? Antworten auf diese Fragen und viele mehr werden auf dem 53. Kongress der Deutschen Gesellschaft für Psychologie präsentiert. Der Kongress startet am Montag, dem 16. September 2024 an der Universität Wien. Der Kongress findet im teilhybriden Format statt, für ausgewählte Veranstaltungen ist auch eine Online-Akkreditierung möglich.

Ähnliche Artikel

53. DGPs-Kongress startet in 10 Tagen: Psychologie in Zeiten gesellschaftlicher Herausforderungen

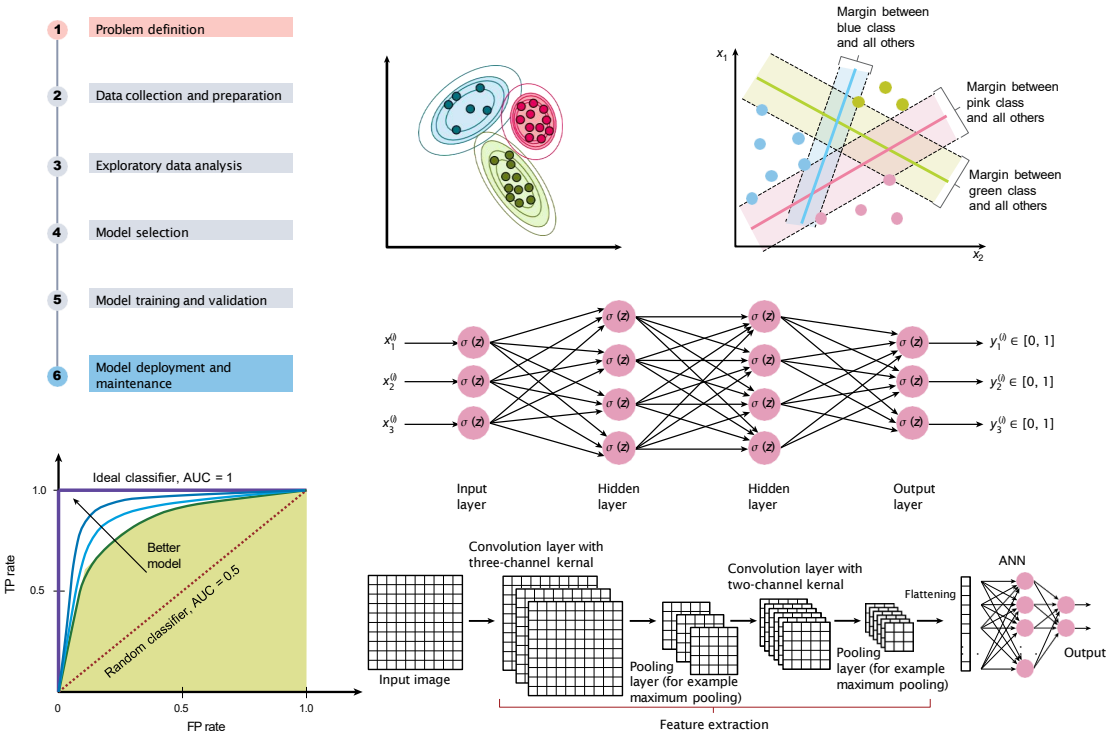
DGPs-Kongress 2024 in Wien: Menschen | Mitwelt | Medien

nature mental health
Review article
https://doi.org/10.1038/s44220-024-00367-2

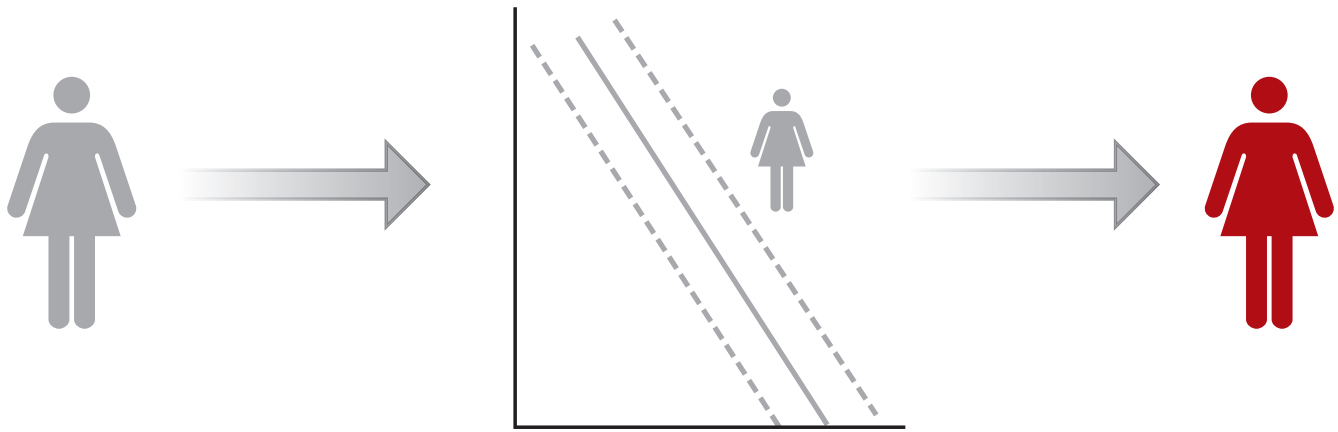
A procedural overview of why, when and how to use machine learning for psychiatry

January 2025

Christopher Lucasius[1], Mai Ali[1], Tanmay Patel[2], Deepa Kundur[1], Peter Szatmari[3,4,5], John Strauss[6] & Marco Battaglia[3,4]

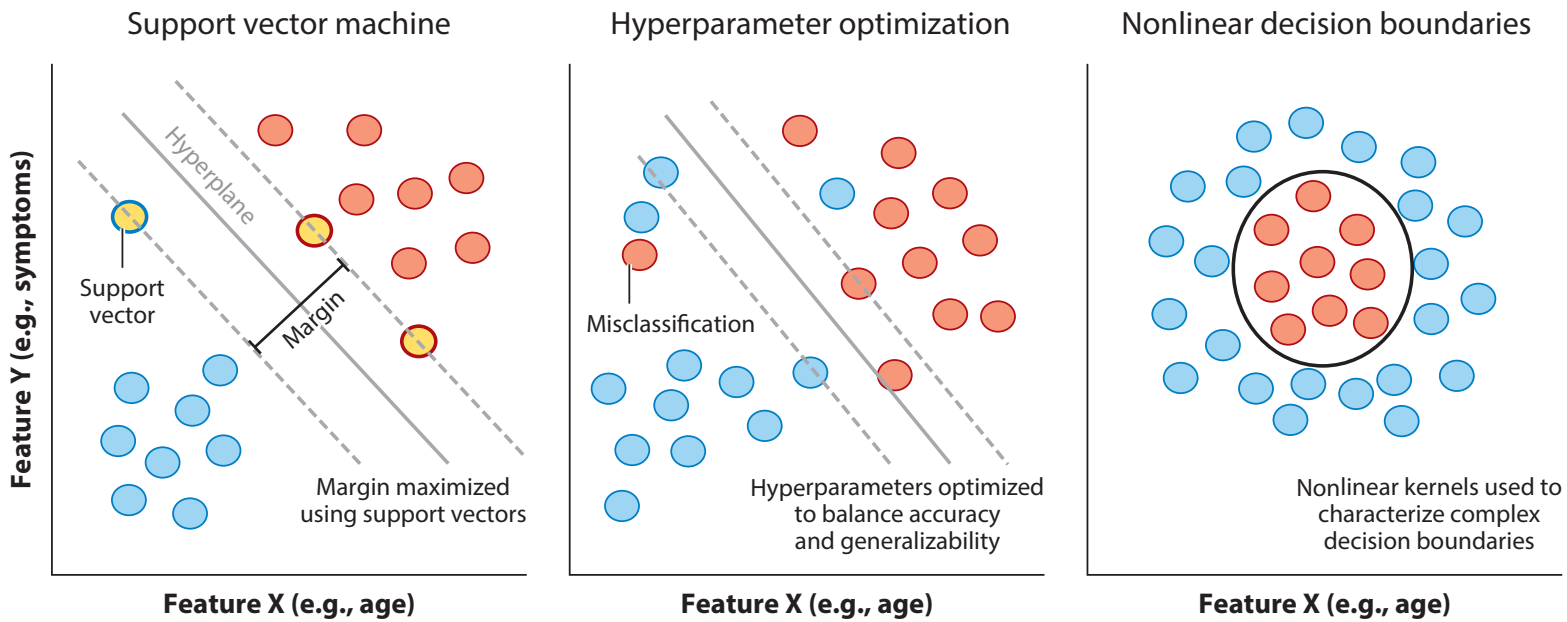


Lucasius et al. (2025)



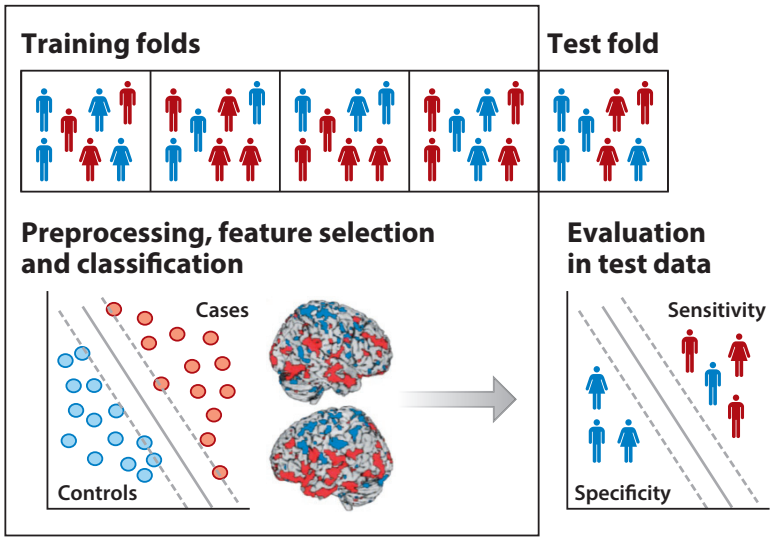
Dwyer et al. (2018) *Annual Reviews of Clinical Psychology*

Rhetorik

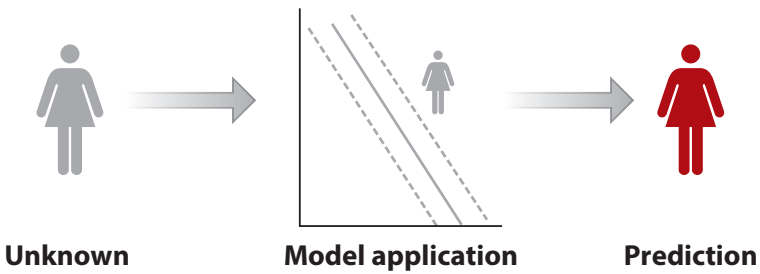


Dwyer et al. (2018) *Annual Reviews of Clinical Psychology*

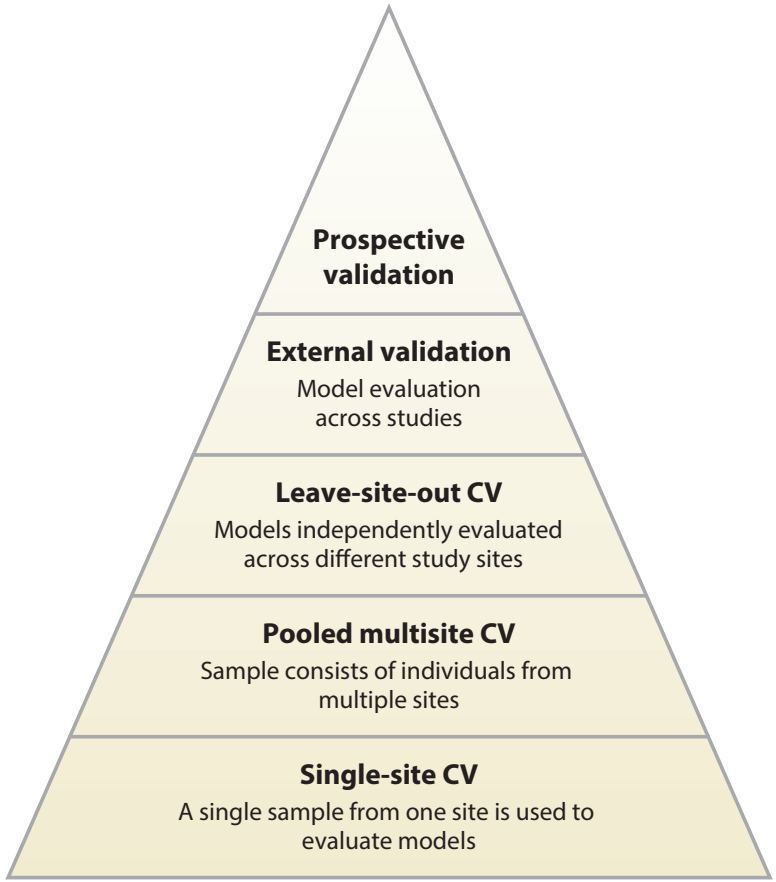
k-fold cross-validation



Prospective validation

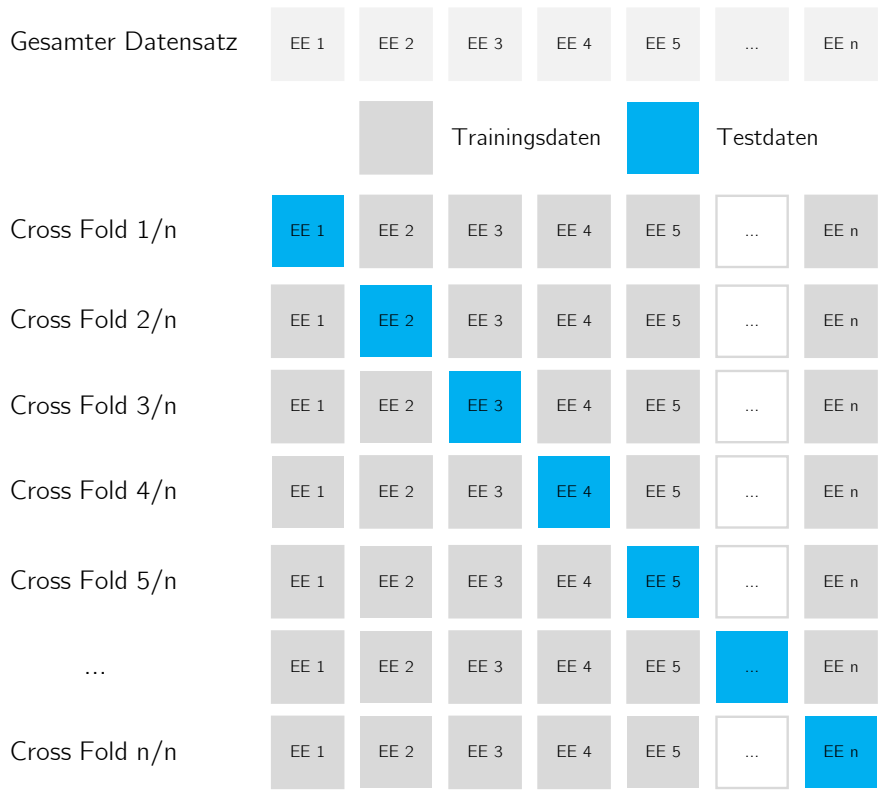


Generalizability hierarchy



Dwyer et al. (2018) *Annual Reviews of Clinical Psychology*

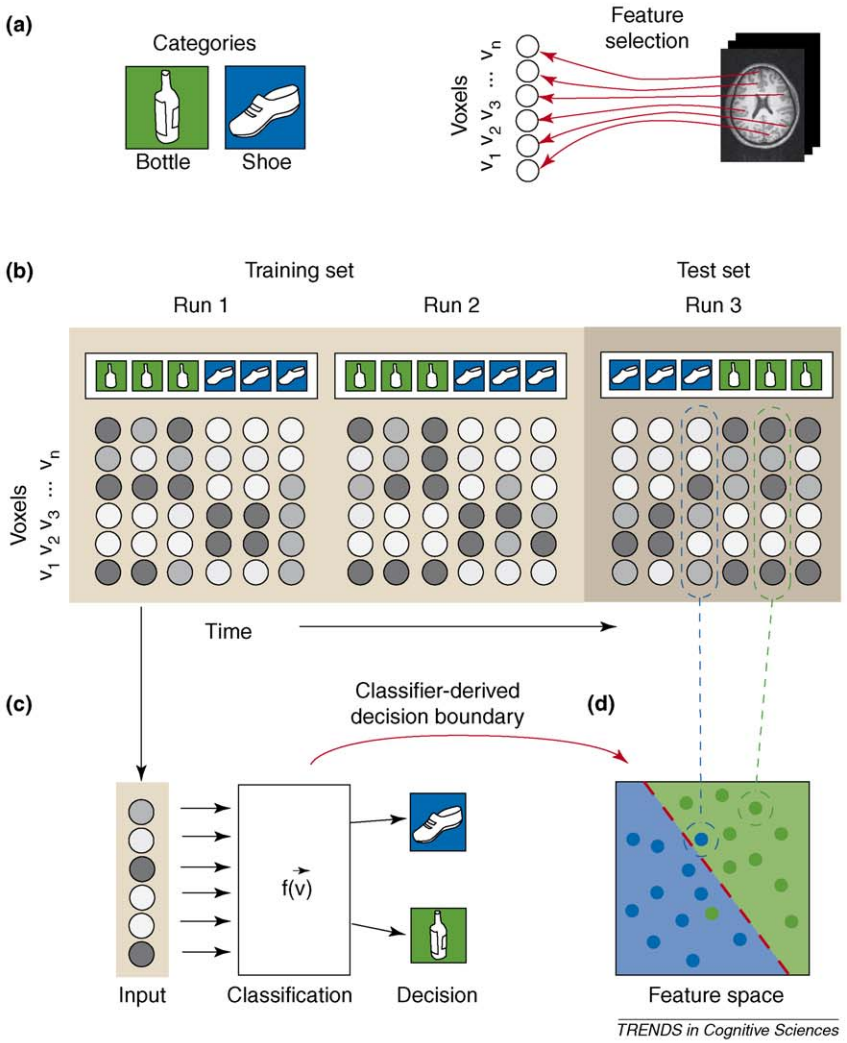
Leave-One-Out-Crossvalidation (LOO-CV)



EE = Experimentelle Einheit (z.B. Proband:in, Studiencenter, ...)

Vgl. Stone (1974)

Rhetorik



Norman et al. (2006) *Trends in Cognitive Science*

Explanatorische Modellierung ⇔ Grundlagenforschung

Bestimmung von $\hat{\phi} \coloneqq \text{argmin} \|\hat{\phi} - \phi\|$

Bestimmung von $\hat{f} \coloneqq \text{argmin}_{f \in F} \|y - f(x)\|, F$ beliebig

Prädiktive Modellierung ⇔ Anwendungsorientierte Forschung

Breiman (2001), Shmueli (2010), Sainani (2014)

Rhetorik

Anwendungsbeispiel

Performanzmetriken

Selbstkontrollfragen

Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach

Nikolaos Koutsouleris, René S Kahn, Adam M Chekroud, Stefan Leucht, Peter Falkai, Thomas Wobrock, Eske M Derks, Wolfgang W Fleischhacker, Alkomiet Hasan

Summary

Background At present, no tools exist to estimate objectively the risk of poor treatment outcomes in patients with first-episode psychosis. Such tools could improve treatment by informing clinical decision-making before the commencement of treatment. We tested whether such a tool could be successfully built and validated using routinely available, patient-reportable information.

Methods By applying machine learning to data from 334 patients in the European First Episode Schizophrenia Trial (EUFEST; International Clinical Trials Registry Platform number, ISRCTN68736636), we developed a tool to predict poor versus good treatment outcome (Global Assessment of Functioning [GAF] score ≥65 *vs* GAF <65, respectively) after 4 weeks and 52 weeks of treatment.To enable the unbiased estimation of the predictive system's generalisability to new patients, we used repeated nested cross-validation to prevent information leaking between patients used for training and validating the models. In pursuit of everyday clinical applicability, we retrained the 4-week outcome predictor with only the top ten predictors of the pooled prediction system and then tested this tool in 108 independent patients with 4-week outcome labels. Discontinuation and readmission to hospital events in patients with predicted poor versus good outcomes were assessed with Kaplan-Meier log-rank analyses, whereas generalised linear mixed-effects models were used to investigate the GAF-based predictions against several clinically meaningful outcome indicators, including treatment adherence, symptom remission, and quality of life.

Koutsouleris et al. (2016)

Anwendungsbeispiel

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Schizophrenia Research 78 (2005) 147–156

SCHIZOPHRENIA RESEARCH

www.elsevier.com/locate/schres

The European First Episode Schizophrenia Trial (EUFEST): Rationale and design of the trial

W. Wolfgang Fleischhacker [a,*], Ireneus P.M. Keet [b], René S. Kahn [b]

EUFEST Steering Committee [1]

[a] *Department of Biological Psychiatry, Medical University Innsbruck, Anichstrasse 35, A-6020 Innsbruck, Austria*
[b] *Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, The Netherlands*

Received 12 May 2005; received in revised form 2 June 2005; accepted 3 June 2005
Available online 1 August 2005

Abstract

Background: Most studies comparing second generation antipsychotics with classical neuroleptics have been conducted in more or less chronic schizophrenia patients. Such studies were usually conducted in highly selected samples, and were generally designed and financed by the manufacturer of the drug tested. These and other facts have stimulated discussions regarding the effectiveness of the new generation of antipsychotics.
Aims: The aim of the European First Episode Schizophrenia Trial (EUFEST) is to compare treatment with amisulpride, quetiapine, olanzapine and ziprasidone to a low dose of haloperidol in an unselected sample of first episode schizophrenia patients with minimal prior exposure to antipsychotics.

Fleischhacker et al. (2005)

Predictor variables and outcome target

Definition of outcome targets, predictor variables, and analysis cohorts

Global Assessment of Functioning (GAF) scores between 61 and 70 have been proposed as thresholds between disease states and at-risk syndromes,[18] indicators of clinically relevant functional impairment,[19] and markers of recovery as part of more complex criteria.[20] Based on these publications, the patients' skewed 1-year follow-up GAF scores, and the resulting low sensitivity of a dimensional prediction model (appendix), we chose a classification approach to differentiate between so-called good (GAF ≥65) and poor (GAF <65) outcome endpoints. An alternative GAF threshold of 60 was assessed by the machine learning methods described below (appendix).

By following a data-driven, multivariate analysis strategy, we deliberately did not pre-select a small set of statistically independent predictor variables. Instead, we set up a rigorously cross-validated learning strategy that autonomously identified the most accurate and parsimonious predictor patterns (figure 1). We initially

Koutsouleris et al. (2016)

Outcome target

Global Assessment of Functioning (GAF)

Fremdeinschätzung der psychischen, sozialen und beruflichen Funktionsfähigkeit

Wertebereich 0–100 (höher = besser)

GAF	Bedeutung
91–100	Keine Symptome, sehr gute Funktionsfähigkeit
81–90	Minimale Symptome, gute Alltagsfunktion
71–80	Vorübergehende Stressreaktionen, leichte Beeinträchtigung
61–70	Leichte Symptome, stabile soziale Beziehungen
51–60	Moderate Symptome oder Funktionsschwierigkeiten
41–50	Schwere Symptome oder deutliche soziale/berufliche Probleme
31–40	Beeinträchtigte Realitätsprüfung oder starke Funktionsstörung
21–30	Verhalten durch Psychose geprägt, kaum funktionsfähig
11–20	Gefahr für sich/andere, massive Beeinträchtigungen
1–10	Suizid-/Fremdgefährdung, minimale Selbstfürsorge
0	Keine ausreichenden Informationen

Wikipedia

Predictor variables

Sociodemographic variables
1 Sex
2 Age
3 Education
4 Weight
5 BMI
6 ECG abnormal
7 Diastolic blood pressure
8 Systolic blood pressure
9 Pulse
10 A023 'Marital status patient: married'
11 A030 'Current occupation patient'
12 A032 'Previous occupation patient'
13 A034 'Highest level occupation father'
14 A034 'Highest level occupation mother'
15 A037 Highest degree patient'
16 A038 'Educational problems'
17 A039 'Education father'
18 A040 'Education mother'
19 A041 'Living alone'
20 A059 'Type of dwelling'
21 A063 'Living environment'

Medication arms
22 Haloperidol treatment
23 Olanzapine treatment
24 Quetiapine treatment
25 Amisulpride treatment
26 Ziprasidone treatment

MINI_Plus interview
27 Former substance abuse
28 Persistent substance abuse
29 Disorganized schizophrenia
30 Catatonic schizophrenia
31 Paranoid schizophrenia
32 Schizophreniform disorder
33 Residual state
34 Schizoaffective disorder
35 Undifferentiated schizophrenia
36 Major depressive episode Current
37 Major depressive episode Recurrent
38 Substance induced mood disorder Past
39 Major depressive episode with melancholic features Current
40 Dysthymia Current
41 Dysthymia Past
42 Suicidality Current
43 Hypomanic episode Past
44 Substance induced manic episode Past
45 Panic disorder Current past month
46 Panic disorder Lifetime
47 Agoraphobia Lifetime
48 Social phobia
49 Specific phobia
50 Obsessive-compulsive disorder
51 Alcohol dependence Past 6 months
52 Alcohol dependence Lifetime
53 Alcohol abuse Past 12 months
54 Alcohol abuse Lifetime
55 Substance dependence Past 12 months
56 Substance dependence Lifetime
57 Substance abuse Past 12 months
58 Substance abuse Lifetime
59 Psychotic disorder Lifetime
60 Mood disorder with psychotic features
61 Schizophrenia Current
62 Schizophrenia Lifetime
63 Schizoaffective disorder Current
64 Schizoaffective disorder Lifetime
65 Schizophreniform disorder Current
66 Schizophreniform disorder Lifetime
67 Substance induced psychotic disorder Lifetime

Camberwell Assessment of Needs
68 CAN_001 Accommodation user Need
69 CAN_007 Accommodation staff Need
70 CAN_011 Food user Need
71 CAN_017 Food staff Need
72 CAN_032 Self-care user Need
73 CAN_038 Self-care staff Need
74 CAN_042 Day time activities user Need
75 CAN_048 Day time activities staff Need
76 CAN_052 Physical Health user Need
77 CAN_058 Physical Health staff Need
78 CAN_062 Psychotic symptoms user Need
79 CAN_068 Psychotic symptoms staff Need
80 CAN_069 Psychotic symptoms staff Informal help given
81 CAN_070 Psychotic symptoms staff Formal help given
82 CAN_071 Psychotic symptoms staff Formal help needed
83 CAN_072 Information user Need
84 CAN_078 Information staff Need
85 CAN_082 Psychological distress user Need
86 CAN_088 Psychological distress staff Need
87 CAN_092 Safety to self user Need
88 CAN_098 Safety to self staff Need
89 CAN_102 Safety to others user Need
90 CAN_108 Safety to others staff Need
91 CAN_118 Alcohol staff Need
92 CAN_122 Drugs user Need
93 CAN_128 Drugs staff Need
94 CAN_132 Company user Need
95 CAN_138 Company staff Need
96 CAN_142 Intimate relationships user Need
97 CAN_149 Intimate relationships staff Need
98 CAN_153 Sexual expression user Need
99 CAN_159 Sexual expression staff Need
100 CAN_174 Education user Need
101 CAN_180 Education staff Need
102 CAN_184 Telephone user Need
103 CAN_194 Transport user Need
104 CAN_200 Transport staff Need
105 CAN_204 Money user Need
106 CAN_210 Money staff Need
107 CAN_214 Benefits user Need
108 CAN_220 Benefits staff Need
109 CAN Summary score (No & Low Need)
110 CAN Summary score (Moderate Need)
111 CAN Summary score (Unmet Need)
112 CGI
113 GAF
114 CDSS

Positive and Negative Symptoms Scale
115 PANSS P1 Delusions
116 PANSS P2 Conceptual disorganization
117 PANSS P3 Hallucinations
118 PANSS P4 Hyperactivity
119 PANSS P5 Grandiosity
120 PANSS P6 Suspiciousness/persecution
121 PANSS P7 Hostility
122 PANSS N1 Blunted affect
123 PANSS N2 Emotional withdrawal
124 PANSS N3 Poor rapport
125 PANSS N4 Passive/apathetic social withdrawal
126 PANSS N5 Difficulty in abstract thinking
127 PANSS N6 Lack of spontaneity and flow of conversation
128 PANSS N7 Stereotyped thinking
129 PANSS G1 Somatic concern
130 PANSS G2 Anxiety
131 PANSS G3 Guilt feelings
132 PANSS G4 Tension
133 PANSS G5 Mannerisms and posturing
134 PANSS G6 Depression
135 PANSS G7 Motor retardation
136 PANSS G8 Uncooperativeness
137 PANSS G9 Unusual thought content
138 PANSS G10 Disorientation
139 PANSS G11 Poor attention
140 PANSS G12 Lack of judgment and insight
141 PANSS G13 Disturbance of volition
142 PANSS G14 Poor impulse control
143 PANSS G15 Preoccupation
144 PANSS G16 Active social avoidance
145 PANSS Positive score
146 PANSS Negative score
147 PANSS General score
148 PANSS Total score

Neuropsychological test battery
149 TMT-A Time
150 TMT-A Errors
151 TMT-A pop
152 TMT-B Time
153 TMT-B Errors
154 TMT-B Pop
155 TMT time (B-A)
156 TMT errors (B-A)
157 Num_digits_Raw-Correct
158 Num_digits_Raw-Skipped
159 Num_digits_Raw-Wrong
160 Num_pegs_Placed-Correct-Dom-Hand
161 Num_pegs_Placed-Correct-NonDom-Hand
162 Num_pegs_Placed-Correct-Both-Hands
163 RAVLT_Tr1_Num_Correct_Responses
164 RAVLT_Tr1_Num_Intrusions
165 RAVLT_Tr1_Num_Repeats
166 RAVLT_Tr2_Num_Correct_Responses
167 RAVLT_Tr2_Num_Intrusions
168 RAVLT_Tr2_Num_Repeats
169 RAVLT_Tr3_Num_Correct_Responses
170 RAVLT_Tr3_Num_Intrusions
171 RAVLT_Tr3_Num_Repeats
172 RAVLT_Tr4_Num_Correct_Responses
173 RAVLT_Tr4_Num_Intrusions
174 RAVLT_Tr4_Num_Repeats
175 RAVLT_Tr5_Num_Correct_Responses
176 RAVLT_Tr5_Num_Intrusions
177 RAVLT_Tr5_Num_Repeats
178 RAVLT_Tr1-5_Num_Correct_Responses
179 RAVLT_Tr1-5_Num_Intrusions
180 RAVLT_Tr1-5_Num_Repeats
181 RAVLT_DR
182 RAVLT_DG5-7
183 RAVLT_DR_Num_Intrusions
184 RAVLT_DR_Num_Repeats

St. Hans Rating Scale for Extrapyramidal Symptoms
185 Akathisia score
186 Dystonia score
187 Parkinsonism score
188 Passive-Dyskinesia score
189 Active-Dyskinesia score

Koutsouleris et al. (2016)

Anwendungsbeispiel

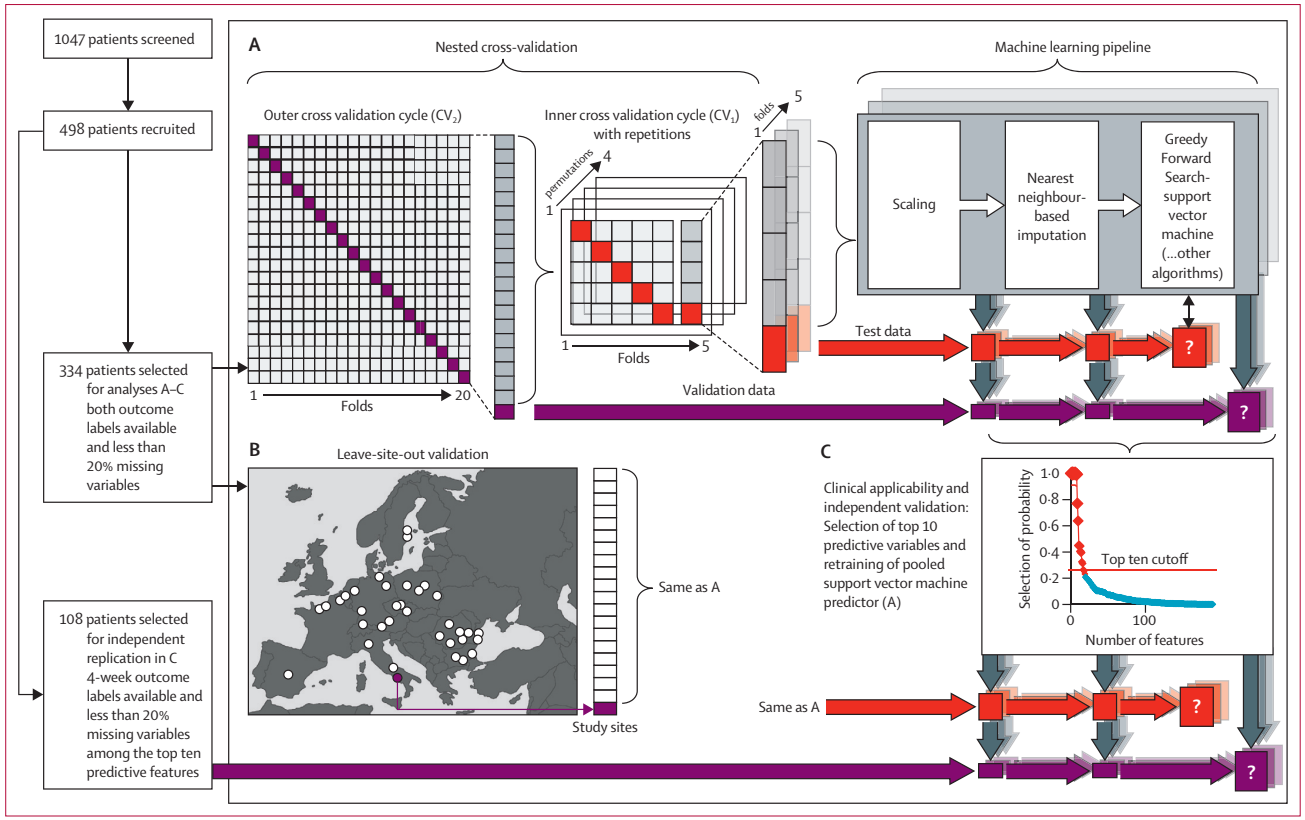


***Figure 1*: Machine learning analysis design**

(A) We used nested, repeated cross-validation to train and validate the 4-week and 52-week outcome predictors, consisting of an outer 20-fold cross-validation cycle (CV_2), which provided validation participants for computing an unbiased estimate of predictor generalisability to new patients, and an inner 4x5-fold cross-validation cycle (CV_1), which delivered training participants to the multivariate pattern analysis pipeline as well as test participants for feature and parameter optimisation. (B) Generalisability of predictors to new sites was assessed by iteratively leaving out once each of 44 the EUFEST study sites included in the analysis, training the predictive system as in A and applying the trained predictor to the left-out site. (C) Using the top ten predictive features of the pooled 4-week outcome classifier, the prediction system was retrained and then applied to 108 patients, who were excluded from the discovery and cross-validation analyses.

Koutsouleris et al. (2016)

Anwendungsbeispiel

Findings The generalisability of our outcome predictions were estimated with cross-validation (test-fold balanced accuracy [BAC] of 75·0% for 4-week outcomes and 73·8% for and 52-week outcomes), and leave-site-out validation across 44 European sites (BAC of 72·1% for 4-week outcomes and 71·1% for 52-week outcomes). We identified a smaller group of ten predictors still providing a BAC of 71·7% in 108 patients never used for model discovery. Unemployment, poor education, functional deficits, and unmet psychosocial needs predicted both endpoints, whereas previous depressive episodes, male sex, and suicidality additionally predicted poor 1-year outcomes. 52-week predictions identified patients at risk for symptom persistence, non-adherence to treatment, readmission to hospital and poor quality of life. Specifically among these patients, amisulpride and olanzapine showed superior efficacy versus haloperidol, quetiapine, and ziprasidone.

Interpretation Our results suggest that prognostic models operating on brief, patient-reportable pre-treatment data might provide useful insight into individualised outcome trajectories, optimising treatment selection, and targeted clinical trial designs. To embed these tools into real-world care, replication is needed in external first-episode samples with overlapping variables, which are not available in the field at present.

Koutsouleris et al. (2016)

Anwendungsbeispiel

	True negatives	True positives	False positives	False negatives	Sensitivity	Specificity	Balanced accuracy	Positive predictive value	Negative predictive value	Prognostic summary index	Positive likelihood ratio	Diagnostic odds ratio	Number-needed-to-diagnose	κ*
4-week outcome predictor: pooled cross-validation analyses														
RBF-SVM	165	84	26	59	73·7%	76·4%	75·0%	86·4%	58·7%	44·1%	3·1	9·7	2·00	..
Linear SVM	170	75	35	54	75·9%	68·2%	72·0%	82·9%	58·1%	41·1%	2·4	5·7	2·27	0·724
Univariate LOGREG	197	55	55	27	88·0%	50·0%	69·0%	78·2%	67·1%	45·2%	1·8	3·1	2·64	0·541
Multivariate LOGREG	197	59	51	27	88·0%	53·6%	70·8%	79·4%	68·6%	48·0%	1·9	3·6	2·40	0·589
Decision Trees	197	54	56	27	88·0%	49·1%	68·5%	77·9%	66·7%	44·5%	1·7	3·0	2·70	0·302
4-week outcome predictor: leave-site-out analyses														
RBF-SVM (10% of top-ranked features)	147	81	29	77	65·6%	73·6%	69·6%	83·5%	51·3%	34·8%	2·5	6·2	2·55	0·814
RBF-SVM (20%)	152	80	30	72	67·9%	72·7%	70·3%	84·4%	54·7%	39·1%	2·5	6·2	2·46	0·837
RBF-SVM (30%)	151	82	28	73	67·4%	74·5%	71·0%	84·0%	54·8%	38·9%	2·6	7·0	2·38	0·807
RBF-SVM (40%)	158	80	30	66	70·5%	72·7%	71·6%	84·4%	52·9%	37·3%	2·6	6·7	2·31	0·837
RBF-SVM (50%)	157	81	29	67	70·1%	73·6%	71·9%	83·5%	52·6%	36·1%	2·7	7·1	2·29	0·873
RBF-SVM (All selected variables)	156	82	28	68	69·6%	74·5%	72·1%	84·8%	54·7%	39·4%	2·7	7·5	2·26	0·909
4-week outcome predictor: condensed battery (top ten features)														
RBF-SVM in 108 new patients	64	13	5	26	71·1%	72·2%	71·7%	92·8%	33·3%	26·1%	2·6	6·6	2·31	..
52-week outcome predictor: pooled cross-validation analyses														
RBF-SVM	52	207	49	26	66·7%	80·9%	73·8%	51·5%	88·8%	40·3%	3·5	12·1	2·10	..
Linear SVM	49	189	67	29	62·8%	73·8%	68·3%	42·2%	87·7%	28·9%	2·4	5·8	2·73	0·734
Univariate LOGREG	13	241	15	65	16·7%	94·1%	55·4%	46·4%	78·8%	25·2%	2·8	8·1	9·25	0·282
Multivariate LOGREG	18	245	11	60	23·1%	95·7%	59·4%	62·1%	80·3%	42·4%	5·4	28·8	5·32	0·331
Decision trees	27	234	21	51	34·6%	91·8%	63·2%	56·3%	82·1%	38·4%	4·2	17·7	3·79	0·429
52-week outcome predictors: leave-site-out analyses														
RBF-SVM (10% of top-ranked features)	49	186	70	29	62·8%	72·7%	67·7%	41·2%	86·5%	27·7%	2·3	5·3	2·82	0·649
RBF-SVM (20%)	52	193	63	26	66·7%	75·4%	71·0%	46·0%	88·2%	34·3%	2·7	7·3	2·38	0·768
RBF-SVM (30%)	51	189	67	27	65·4%	73·8%	69·6%	45·3%	88·5%	33·8%	2·5	6·2	2·55	0·777
RBF-SVM (40%)	53	192	64	25	67·9%	75·0%	71·5%	43·2%	87·5%	30·7%	2·7	7·4	2·33	0·810
RBF-SVM (50%)	52	195	61	26	66·7%	76·2%	71·4%	45·2%	88·1%	33·3%	2·8	7·8	2·33	0·808
RBF-SVM (All selected variables)	50	200	56	28	64·1%	78·1%	71·1%	47·2%	87·7%	34·9%	2·9	8·6	2·37	0·797

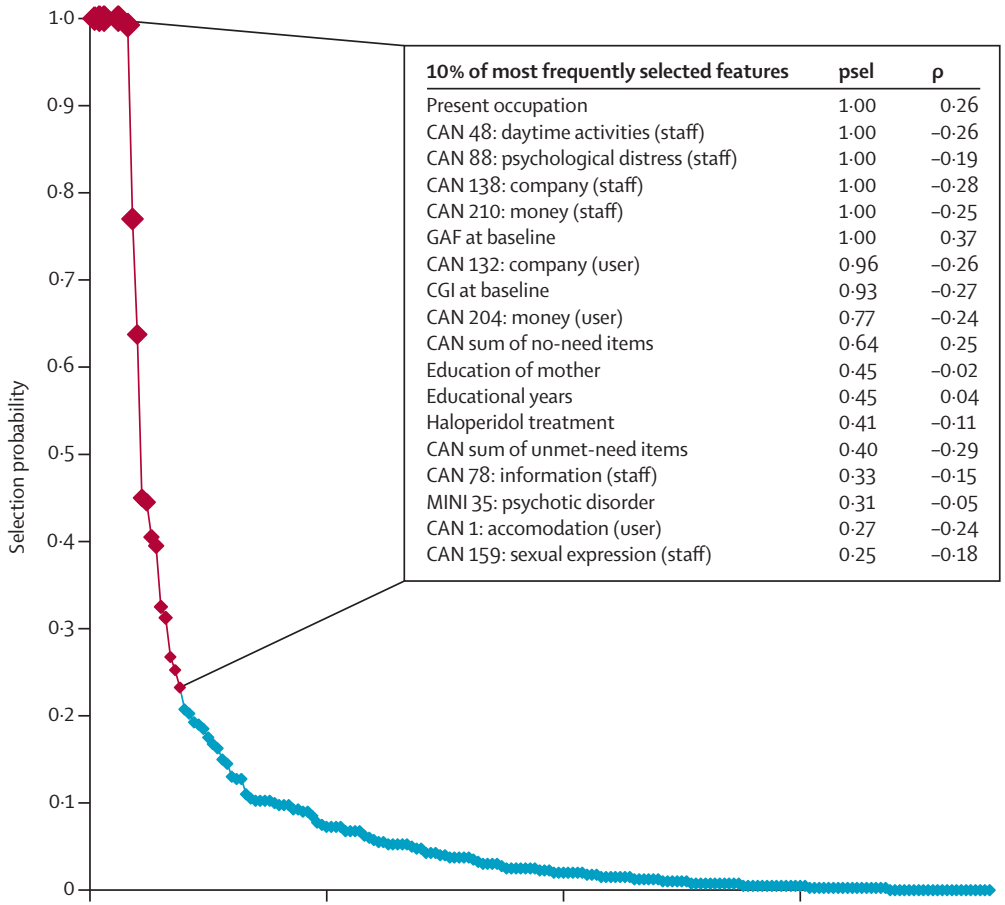
Based on the established sensitivity and specificity nomenclature, we label predicted poor outcomes as positive predictions and predicted good outcomes as negative predictions, hence sensitivity measures the classifiers' capacity to correctly identify patients with poor outcomes as such. Methodological validation of SVM was done against univariate and L2-regularised multivariate logistic regression models as well as decision tree ensembles in both 4-week and 52-week pooled prediction analyses. Prediction performance was further validated in leave-site-out analyses without clinical variable confinement (all selected variables) as well as with different degrees of top-variable selection (10–50%). An additional validation step was done for the pooled 4-week SVM predictor, which was trained on the top ten most predictive features (figure 2) and then applied to a cohort of 108 EUFEST patients never included in any analysis procedure.

SVM=support Vector Machine. LOGREG=logistic regression. RBF=non-linear radial basis function kernel. SVM=support vector machine. *κ=Cohen's kappa measuring predictor agreement between the Pooled SVM and the respective predictor.

Table 2: **Validated predictive performance of classifiers trained on the 4-week and 52-week outcome endpoints**

Koutsouleris et al. (2016)

Anwendungsbeispiel



10% of most frequently selected features	psel	ρ
Present occupation	1·00	0·26
CAN 48: daytime activities (staff)	1·00	−0·26
CAN 88: psychological distress (staff)	1·00	−0·19
CAN 138: company (staff)	1·00	−0·28
CAN 210: money (staff)	1·00	−0·25
GAF at baseline	1·00	0·37
CAN 132: company (user)	0·96	−0·26
CGI at baseline	0·93	−0·27
CAN 204: money (user)	0·77	−0·24
CAN sum of no-need items	0·64	0·25
Education of mother	0·45	−0·02
Educational years	0·45	0·04
Haloperidol treatment	0·41	−0·11
CAN sum of unmet-need items	0·40	−0·29
CAN 78: information (staff)	0·33	−0·15
MINI 35: psychotic disorder	0·31	−0·05
CAN 1: accomodation (user)	0·27	−0·24
CAN 159: sexual expression (staff)	0·25	−0·18

Koutsouleris et al. (2016)

Anwendungsbeispiel

Research in context

Evidence before this study

We systematically searched PubMed for all papers published in English up to April 7, 2016, using the following search terms "(prediction) AND (psychosis OR schizophrenia) AND (machine learning OR multivariate pattern analysis OR support vector machine)". We identified 31 unique records, none of which presented a thoroughly validated approach to predict the treatment outcomes of individual patients with first-episode psychosis using broadly accessible clinical baseline data. Although many group-level sociodemographic, clinical, and neurocognitive variables are known to be moderately associated with therapeutic outcomes, it is unclear how and which of these variables are effectively combined into prognostic models so that accurate and generalisable outcome predictions can be generated on a case-by-case level. Furthermore, how these predictions could inform therapeutic decision making so that the right patients receive the most appropriate treatments is unknown.

Added value of this study

This the first study to show accurate and generalisable individual-patient outcome predictions after 4 and 52 weeks of treatment for first-episode psychosis. Our machine learning models used only clinical baseline information, particularly psychosocial, sociodemographic and psychometric variables, and as few as ten questions. Notably, the resulting predictions tracked an array of clinically meaningful outcome indicators, such as treatment adherence, readmission to hospital, functioning, symptom burden, and quality of life, thus approximating the more complex phenotype of recovery from first-episode psychosis. The present approach might provide a more than 40% gain in correctly predicting outcomes across geographically distinct health-care settings and patient populations. Importantly, our models revealed that the efficacy differences previously found among the five European First Episode Schizophrenia Trial antipsychotic treatments, and those reported in recent meta-analytic comparisons, might be driven by the group of patients with poor prognoses, as predicted by our analytical tools.

Implications of all the available evidence

Our study exemplifies methodological advances crucial to the development of reliable outcome prediction models in the psychosis field. In this context, the evidence generated by our study strongly suggests that measuring individual risk by prognostic modelling could inform everyday clinical care as well as new clinical trials designs in the era of digital mental health.

Koutsouleris et al. (2016)

Rhetorik

Anwendungsbeispiel

Performanzmetriken

Selbstkontrollfragen

Binärer Klassifikationstrainingdatensatz

Ein *binärer Klassifikationsdatensatz*

$$\mathcal{D} := \{(x_1, y_1), ..., (x_n, y_n)\} = \{(x_i, y_i)\}_{i=1}^{n} \tag{1}$$

ist eine Menge von n *Trainingsdatenpunkten*

$$(x_i, y_i) \text{ mit } x_i \in \mathbb{R}^m \text{ und } y_i \in \{0, 1\} \text{ for } i = 1, ..., n, \tag{2}$$

wobei x_i m-*dimensionaler Featurevektor* und y_i *Label* genannt wird. Üblicherweise werden die Trainingsdatenpunkte dabei als unabhängige und identische Realisierungen eines $m+1$-dimensionalen Zufallsvektors $\zeta := (\xi, \upsilon)$ verstanden.

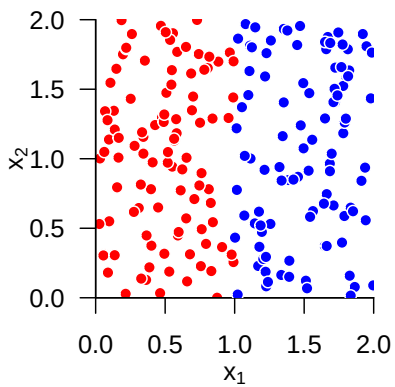
Bemerkungen

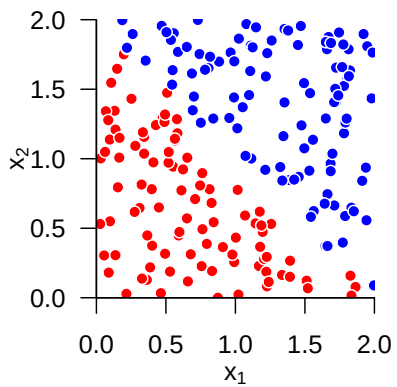
- $y_i \in \{0, 1\}$ bezeichnet die Klassenzugehörigkeit des Featurevektors $x_i \in \mathbb{R}^m$.
- Ein Beispiel für y_i ist "GAF $<$ 65" (0)" vs. "GAF $\geq$ 65" (1).
- Beispiele für die m Komponenten der x_i sind Testscores, soziodemographische Daten, ...

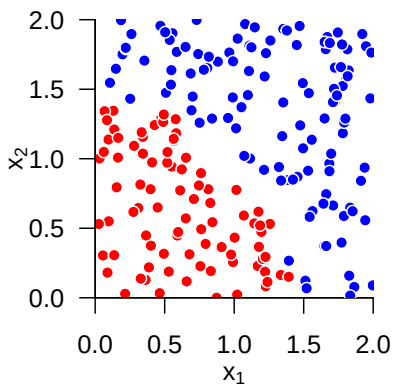
Beispiele bivariater Featureplotszenarien

$$x_i \in \mathbb{R}^2, y_i \in \{0, 1\}, i = 1, ..., n, \bullet\ y_i = 0, \quad \bullet\ y_i = 1$$

$$x_{i_1} > 1 \Leftrightarrow y_i = 1 \qquad x_{i_1} + x_{i_2} > 2 \Leftrightarrow y_i = 1 \qquad x_{i_1}^2 + x_{i_2}^2 > 2 \Leftrightarrow y_i = 1$$







Sprache der prädiktiven Modellierung

Daten	Trainingsdaten und Testdaten
Statistisches Modell	Modell, Algorithmus, “Die künstliche Intelligenz”
Schätzen von Parametern	Trainieren des Modells, Lernen, Supervised Learning

Workflow der prädiktiven Modellierung

Featureselektion

- Auswahl von möglichen prädiktiven Variablen
- Dimensionreduktion zur Verringerung des Curse of Dimensionality

Kreuzvalidierung

- Wiederholtes Trainieren und Testen eines Modells an einem Datensatz
- Einsatz zur Modelloptimierung
- Einsatz zur Messung der probabilistischen Assoziation von Featuren und Labeln

Konfusionsmatrix bei LOO-CV mit binärem Label für (x_i, y_i)

Fall \ Prädiktion	$f(x_i) = 0$	$f(x_i) = 1$	
$y_i = 0$	True Negative TN	False Positive FP	Negative $N := TN + FP$
$y_i = 1$	False Negative FN	True Positive TP	Positive $P := FN + TP$
	Predicted Negative $PN := TN + FN$	Predicted Positive $PP := TP + FP$	Total $T := N + P$

True Positive = Hit, False Positive = False alarm, True Negative = Correct rejection, False Negative = Miss

Prädiktionsgüte

$$\text{Accuracy (ACC)} = \frac{\text{Anzahl richtiger Prädiktionen}}{\text{Anzahl aller Prädiktionen}} = \frac{TN + TP}{T} \tag{3}$$

Prädiktionsraten mit Bezug zu tatsächlichen Fällen

True Positive Rate (aka Hit Rate, Sensitivity, Detection probability, Recall, Power)

$$\text{TPR} = \frac{\text{Anzahl richtiger Positivprädiktionen}}{\text{Anzahl positiver Fälle}} = \frac{\text{TP}}{\text{P}} \tag{4}$$

False Positive Rate (aka Type I Error Rate, False Alarm Probability, Fall-Out)

$$\text{FPR} = \frac{\text{Anzahl falscher Positivprädiktionen}}{\text{Anzahl negativer Fälle}} = \frac{\text{FP}}{\text{N}} \tag{5}$$

True Negative Rate (aka Specificity, Selectivity)

$$\text{TNR} = \frac{\text{Anzahl richtiger Negativprädiktionen}}{\text{Anzahl negativer Fälle}} = \frac{\text{TN}}{\text{N}} \tag{6}$$

False Negative Rate (aka Type II Error Rate, Miss Rate)

$$\text{FNR} = \frac{\text{Anzahl falscher Negativprädiktionen}}{\text{Anzahl positiver Fälle}} = \frac{\text{FN}}{\text{P}} \tag{7}$$

Beispiel (Koutsouleris et al. (2016), Table 1, 4-week outcome prediction, RBF-SVM)

		Prädiktion: $f(x_i) = 0$	Prädiktion: $f(x_i) = 1$	
Fall	$y_i = 0$	True Negative $\mathrm{TN} = 165$	False Positive $\mathrm{FP} = 26$	Negative $\mathrm{N} = 191$
	$y_i = 1$	False Negative $\mathrm{FN} = 59$	True Positive $\mathrm{TP} = 84$	Positive $\mathrm{P} = 143$
		Predicted Negative $\mathrm{PN} = 224$	Predicted Positive $\mathrm{PP} = 110$	Total $\mathrm{T} = 334$

Prädiktionsgüte

$$\text{Accuracy (ACC)} = \frac{\text{Anzahl richtiger Prädiktionen}}{\text{Anzahl aller Prädiktionen}} = \frac{\mathrm{TN} + \mathrm{TP}}{\mathrm{T}} = \frac{165 + 84}{334} = 74.6\% \qquad (8)$$

Performanzmetriken

Beispiel (Koutsouleris et al. (2016), Table 1, 4-week outcome prediction, RBF-SVM)

True Positive Rate (aka Hit Rate, **Sensitivity**, Detection probability, Recall, Power)

$$\text{TPR} = \frac{\text{Anzahl richtiger Positivprädiktionen}}{\text{Anzahl positiver Fälle}} = \frac{\text{TP}}{\text{P}} = \frac{84}{143} = 58.7\% \tag{9}$$

False Positive Rate (aka Type I Error Rate, False Alarm Probability, Fall-Out)

$$\text{FPR} = \frac{\text{Anzahl falscher Positivprädiktionen}}{\text{Anzahl negativer Fälle}} = \frac{\text{FP}}{\text{N}} = \frac{26}{191} = 13.6\% \tag{10}$$

True Negative Rate (aka **Specificity**, Selectivity)

$$\text{TNR} = \frac{\text{Anzahl richtiger Negativprädiktionen}}{\text{Anzahl negativer Fälle}} = \frac{\text{TN}}{\text{N}} = \frac{165}{191} = 86.4\% \tag{11}$$

False Negative Rate (aka Type II Error Rate, Miss Rate)

$$\text{FNR} = \frac{\text{Anzahl falscher Negativprädiktionen}}{\text{Anzahl positiver Fälle}} = \frac{\text{FN}}{\text{P}} = \frac{59}{143} = 41.3\% \tag{12}$$

Prädiktionsraten mit Bezug zu Prädiktionsanzahlen

Positive Predictive Value / Precision

$$\text{PPV} = \frac{\text{Anzahl richtiger Positivprädiktionen}}{\text{Anzahl positiver Prädiktionen}} = \frac{\text{TP}}{\text{PP}} \tag{13}$$

False Discovery Rate

$$\text{FDR} = \frac{\text{Anzahl falscher Positivprädiktionen}}{\text{Anzahl positiver Prädiktionen}} = \frac{\text{FP}}{\text{PP}} \tag{14}$$

Negative Predictive Value

$$\text{NPV} = \frac{\text{Anzahl richtiger Negativprädiktionen}}{\text{Anzahl negativer Prädiktionen}} = \frac{\text{TN}}{\text{PN}} \tag{15}$$

False Omission Rate

$$\text{NPV} = \frac{\text{Anzahl falscher Negativprädiktionen}}{\text{Anzahl negativer Prädiktionen}} = \frac{\text{FN}}{\text{PN}} \tag{16}$$

Performanzmetriken

Beispiel (Koutsouleris et al. (2016), Table 1, 4-week outcome prediction, RBF-SVM)

Positive Predictive Value / Precision

$$\text{PPV} = \frac{\text{Anzahl richtiger Positivprädiktionen}}{\text{Anzahl positiver Prädiktionen}} = \frac{\text{TP}}{\text{PP}} = \frac{84}{110} = 76.4\% \tag{17}$$

False Discovery Rate

$$\text{FDR} = \frac{\text{Anzahl falscher Positivprädiktionen}}{\text{Anzahl positiver Prädiktionen}} = \frac{\text{FP}}{\text{PP}} = \frac{26}{110} = 23.6\% \tag{18}$$

Negative Predictive Value

$$\text{NPV} = \frac{\text{Anzahl richtiger Negativprädiktionen}}{\text{Anzahl negativer Prädiktionen}} = \frac{\text{TN}}{\text{PN}} = \frac{165}{224} = 73.7\% \tag{19}$$

False Omission Rate

$$\text{NPV} = \frac{\text{Anzahl falscher Negativprädiktionen}}{\text{Anzahl negativer Prädiktionen}} = \frac{\text{FN}}{\text{PN}} = \frac{59}{224} = 26.3\% \tag{20}$$

Balanced Accuracy

Bei ungleichen Verhältnissen von Fallzahlen ist Accuracy (ACC) leicht missverständlich

Seien bspw. $\mathrm{N} = 95$ und $\mathrm{P} = 5$ und alle Prädiktionen negativ. Dann gilt

$$\mathrm{ACC} = \frac{\mathrm{TN} + \mathrm{TP}}{\mathrm{T}} = \frac{95 + 0}{100} = 95\% \tag{21}$$

Die Balanced Accuracy ist der Mittelwert der True Positive Rate und der True Negative Rate

$$\text{Balanced Accuracy (BAC)} = \frac{\mathrm{TPR} + \mathrm{TNR}}{2} \tag{22}$$

Seien $\mathrm{N} = 95$ und $\mathrm{P} = 5$ und alle Prädiktionen negativ. Dann gilt

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}} = \frac{0}{5} = 0\%, \quad \mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{N}} = \frac{95}{95} = 100\%, \quad \mathrm{BAC} = \frac{\mathrm{TPR} + \mathrm{TNR}}{2} = \frac{0\% + 100\%}{2} = 50\%$$

Seien $\mathrm{N} = 95$ und $\mathrm{P} = 5$ und alle Negativprädiktionen und alle Positivprädiktionen richtig. Dann gilt

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}} = \frac{5}{5} = 100\%, \quad \mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{N}} = \frac{95}{95} = 100\%, \quad \mathrm{BAC} = \frac{\mathrm{TPR} + \mathrm{TNR}}{2} = \frac{100\% + 100\%}{2} = 100\%$$

Beispiel (Koutsouleris et al. (2016), Table 1, 4-week outcome prediction, RBF-SVM)

$$\mathrm{BAC} = \frac{\mathrm{TPR} + \mathrm{TNR}}{2} = \frac{58.7\% + 86.4\%}{2} = 72.6\% \tag{23}$$

F1 Score

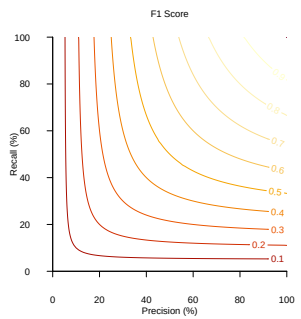
F1 Score ist ein integriertes Maß von Precision und Recall aus der Information Retrieval Literatur

$$\text{Precision} = \frac{\text{TP}}{\text{PP}} = \text{PPV} \qquad \text{Recall} = \frac{\text{TP}}{\text{P}} = \text{TPR} \tag{24}$$

Es gilt

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{25}$$

Der F1 Score ist groß, wenn sowohl Precision als auch Recall groß sind.



Ist Precision oder Recall oder beide gering, so ist der F1 Score niedrig

Beispiel (Koutsouleris et al. (2016), Table 1, 4-week outcome prediction, RBF-SVM)

$$\text{F1} = 2 \cdot \frac{76.4\% \cdot 58.7\%}{76.4\% + 58.7\%} = 66.4\% \tag{26}$$

Rhetorik

Anwendungsbeispiel

Performanzmetriken

Selbstkontrollfragen

Selbstkontrollfragen

1. Erläutern Sie die Rhetorik der Prädiktiven Modellierung.

2. Erläutern Sie Unterschiede und Gemeinsamkeiten der explanatorischen und prädiktiven Modellierung.

3. Erläutern Sie die Idee der Leave-One-Out-Crossvalidation (LOO-CV).

4. Erläutern und definieren Sie die Konfusionsmatrix bei LOO-CV mit binärem Label.

5. Geben Sie die Definitionen von True/False Positive Rate und True/False Negative Rate wieder.

6. Geben Sie die Definitionen von Positive/Negative Prediction Value und False Discovery/Omission Rate wieder.

7. Erläutern und definieren Sie den Begriff der Balanced Accuracy.

Referenzen I

Breiman, L. (2001). Statistical Modeling: The Two Cultures. *Statistical Science*, *16*(3), 199–215. https://www.jstor.org/stable/2676681

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annual Review of Clinical Psychology*, *14*(1), 91–118. https://doi.org/10.1146/annurev-clinpsy-032816-045037

Fleischhacker, W. W., Keet, I. P. M., & Kahn, R. S. (2005). The European First Episode Schizophrenia Trial (EUFEST): Rationale and design of the trial. *Schizophrenia Research*, *78*(2-3), 147–156. https://doi.org/10.1016/j.schres.2005.06.004

Koutsouleris, N., Kahn, R. S., Chekroud, A. M., Leucht, S., Falkai, P., Wobrock, T., Derks, E. M., Fleischhacker, W. W., & Hasan, A. (2016). Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: A machine learning approach. *The Lancet Psychiatry*, *3*(10), 935–946. https://doi.org/10.1016/S2215-0366(16)30171-7

Lucasius, C., Ali, M., Patel, T., Kundur, D., Szatmari, P., Strauss, J., & Battaglia, M. (2025). A procedural overview of why, when and how to use machine learning for psychiatry. *Nature Mental Health*, *3*(1), 8–18. https://doi.org/10.1038/s44220-024-00367-2

Meisenzahl, E., & Sprick, U. (2023). *E-Mental-Health in Psychiatrie und Psychotherapie: Digitale Gesundheitsanwendungen, Online-Therapieprogramme, Videosprechstunden & Co*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-64457-7

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430. https://doi.org/10.1016/j.tics.2006.07.005

Pfannstiel, M. A. (Ed.). (2022). *Künstliche Intelligenz im Gesundheitswesen: Entwicklungen, Beispiele und Perspektiven*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-33597-7

Referenzen II

Sainani, K. L. (2014). Explanatory Versus Predictive Modeling. *PM&R*, *6*(9), 841–844. https://doi.org/10.1016/j.pmrj.2014.08.941

Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, *25*(3). https://doi.org/10.1214/10-STS330

Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, *36*(2), 111–147. https://www.jstor.org/stable/2984809