Additional file 1

Development and internal validation of a CSF-free machine-learning model to distinguish immune checkpoint inhibitor-induced encephalitis from HSV-1 and anti-LGI1 encephalitis

BMC Neurology

Sorama Aoki and Junko Kawakami

Division of Medical and Pharmaceutical Information Science, Faculty of Pharmaceutical Sciences, Tohoku Medical and Pharmaceutical University, Sendai, Japan

Corresponding author: Sorama Aoki (s-aoki@tohoku-mpu.ac.jp)

Contents: Supplementary Methods; Supplementary Tables S1–S2; Supplementary Figures S1–S3

Supplementary Methods

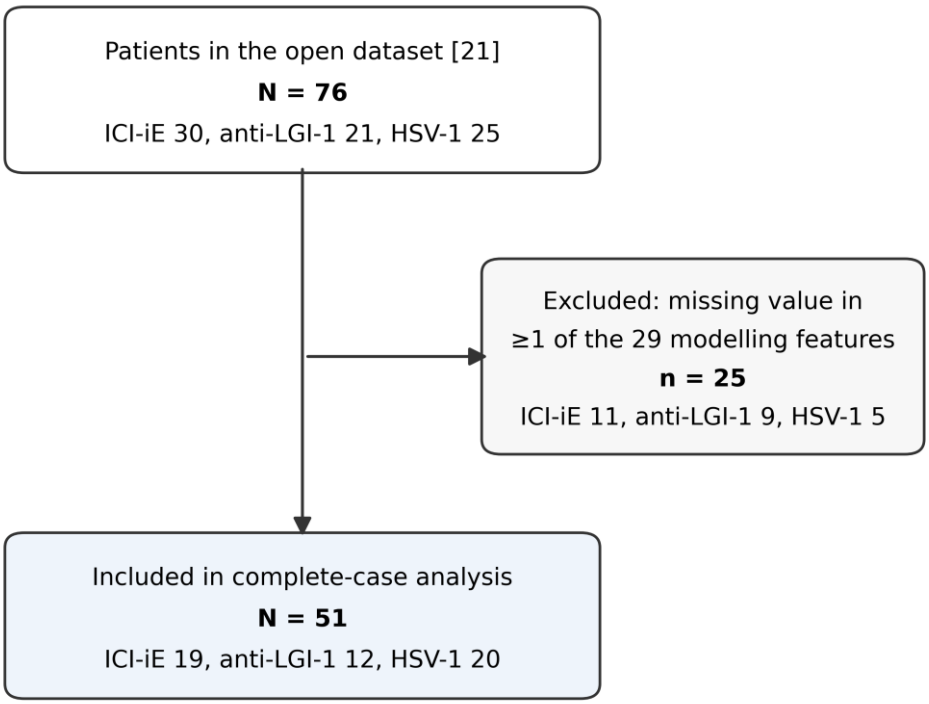
Model development and hyperparameter search. Random forests were fitted with 300 trees and a minimum of 2 samples required to split a node. The inner grid search explored maximum tree depth (3, 5, or unlimited), minimum samples per leaf (1, 2, or 4), the number of features considered at each split ($\sqrt{p}$ or $\log_2 p$), and class weight (none or balanced), using the area under the receiver operating characteristic curve (AUC) as the optimisation criterion.

Nested cross-validation. The outer loop was 5-fold cross-validation repeated 20 times (repeated stratified k-fold). For each outer training portion a separate grid search was performed by inner 5-fold cross-validation, and the model with the selected hyperparameters was used to predict the corresponding outer test portion. Out-of-fold (OOF) predicted probabilities and AUC were obtained for all cases, and discrimination was summarised as the mean, standard deviation, and 2.5–97.5th percentile interval across the 20 repeats. The random seed was fixed at 42 throughout. Calibration, decision-curve analysis, and between-model comparisons used per-case OOF predicted probabilities averaged across repeats.

Threshold determination. In each outer split, the minimum threshold achieving 100% specificity was determined from the OOF predicted probabilities of the inner training portion and applied to the outer test portion. Confusion matrices were then aggregated across all outer splits and all repeats to compute sensitivity, specificity, and positive predictive value on the outer data.

SHAP computation. To define the interpretation target uniquely, a single hyperparameter set was selected for each feature set by grid search with inner 5-fold cross-validation on all data, and SHAP values were computed under that set as the average of 10 rounds of stratified 5-fold cross-validation, ensuring that they were out-of-fold. The selected hyperparameters were 300 trees, a minimum of 1 sample per leaf, and no class weighting in both models, with a maximum depth of 5 and $\log_2 p$ features per split for the with-CSF model and a maximum depth of 3 and $\sqrt{p}$ features per split for the without-CSF model.

Calibration and decision-curve analysis. Calibration was evaluated on the OOF predicted probabilities using the observed proportion within five quantile-based bins (with 95% Wilson confidence intervals), a LOWESS-smoothed calibration curve, the calibration slope and intercept (the coefficient and intercept of a logistic regression of the outcome on logit(p)), and the Brier score. Decision-curve analysis assumed a decision to classify a patient as ICI-iE and manage accordingly, and the 95% confidence interval of net benefit was estimated by bootstrap resampling of cases (500 replicates).



Supplementary Figure S1.

Flow of patients through the complete-case analysis. Of the 76 patients in the source dataset (ICI-iE, n = 30; anti-LGI1 encephalitis, n = 21; HSV-1 encephalitis, n = 25), 25 were excluded because of a missing value in at least one of the 29 modelling features (ICI-iE, n = 11; anti-LGI1 encephalitis, n = 9; HSV-1 encephalitis, n = 5), leaving 51 patients (ICI-iE, n = 19; anti-LGI1 encephalitis, n = 12; HSV-1 encephalitis, n = 20) for analysis. ICI-iE, immune checkpoint inhibitor-induced encephalitis; HSV-1, herpes simplex virus type 1; LGI1, leucine-rich glioma-inactivated 1.

Supplementary Table S1.

Variables recorded in the source dataset but excluded prior to model development owing to missing data. For each variable, the number (percentage) of patients with a missing value is shown by encephalitis group and overall, calculated among all recorded patients before complete-case selection (ICI-iE, n = 30; anti-LGI1 encephalitis, n = 21; HSV-1 encephalitis, n = 25; total, n = 76). Variables were removed when missing values were frequent overall or concentrated within a single disease group, because retaining the latter would have selectively removed that group's patients during complete-case selection and further reduced and unbalanced the already small class sizes; for example, missing values for mnestic deficit, vomiting/nausea, and myoclonus occurred only in the ICI-iE group, whereas missing values for hyponatremia were most frequent in the anti-LGI1 group. CSF, cerebrospinal fluid; EEG, electroencephalography; HSV-1, herpes simplex virus type 1; ICI-iE, immune checkpoint inhibitor-induced encephalitis; LGI1, leucine-rich glioma-inactivated 1.

Variables	Category	ICI-iE (n = 30)	anti-LGI1 (n = 21)	HSV-1 (n = 25)	Total (n = 76)
Procalcitonin at onset	Laboratory (blood)	21 (70%)	19 (90%)	12 (48%)	52 (68%)
Intrathecal immunoglobulin synthesis	CSF	16 (53%)	10 (48%)	9 (36%)	35 (46%)
CSF cell differentiation	CSF	11 (37%)	18 (86%)	5 (20%)	34 (45%)
EEG: diffuse slowing / non-epileptic	EEG	13 (43%)	3 (14%)	9 (36%)	25 (33%)
EEG: epileptic abnormalities	EEG	13 (43%)	3 (14%)	9 (36%)	25 (33%)
Serum sodium at onset	Laboratory (blood)	16 (53%)	7 (33%)	0 (0%)	23 (30%)
Hyponatremia	Laboratory (blood)	3 (10%)	6 (29%)	1 (4%)	10 (13%)
Vomiting / nausea	Symptom/sign	9 (30%)	0 (0%)	0 (0%)	9 (12%)
Myoclonus	Symptom/sign	9 (30%)	0 (0%)	0 (0%)	9 (12%)
Mnestic deficit	Symptom/sign	5 (17%)	0 (0%)	0 (0%)	5 (7%)

Values are the number (percentage) of patients with a missing value for each variable, among all patients in the source dataset before complete-case selection; percentages use the group totals in the header. Variables were excluded when missing values were frequent overall or concentrated within a single group (e.g., missing values for mnestic deficit, vomiting/nausea, and myoclonus occurred only in the ICI-iE group). CSF, cerebrospinal fluid; EEG, electroencephalography.

Supplementary Table S2.

Discrimination of the three feature sets, with pairwise comparisons. Three random forest models — with cerebrospinal fluid (CSF) features (29 features), without CSF features (25), and without CSF and MRI features (23) — were trained and evaluated under settings identical to the main analysis, differing only in the feature set; the last set additionally removes the two MRI features and represents an exploratory sensitivity analysis rather than a proposed clinical model. (A) Discrimination, summarized as the mean (standard deviation) and 2.5–97.5th percentile of the per-repeat pooled out-of-fold AUC across the nested cross-validation (5 outer folds × 20 repeats), together with the AUC of the per-case out-of-fold probabilities averaged across repeats (the curves plotted in Fig. S2). (B) Pairwise differences in AUC (ΔAUC) from the DeLong test applied to the same repeat-averaged probabilities (an approximate paired comparison; a positive ΔAUC favours the first-named set); ΔAUC therefore corresponds to the difference between the repeat-averaged AUCs in panel A, not between the mean per-repeat AUCs. AUC, area under the receiver operating characteristic curve; CSF, cerebrospinal fluid; MRI, magnetic resonance imaging.

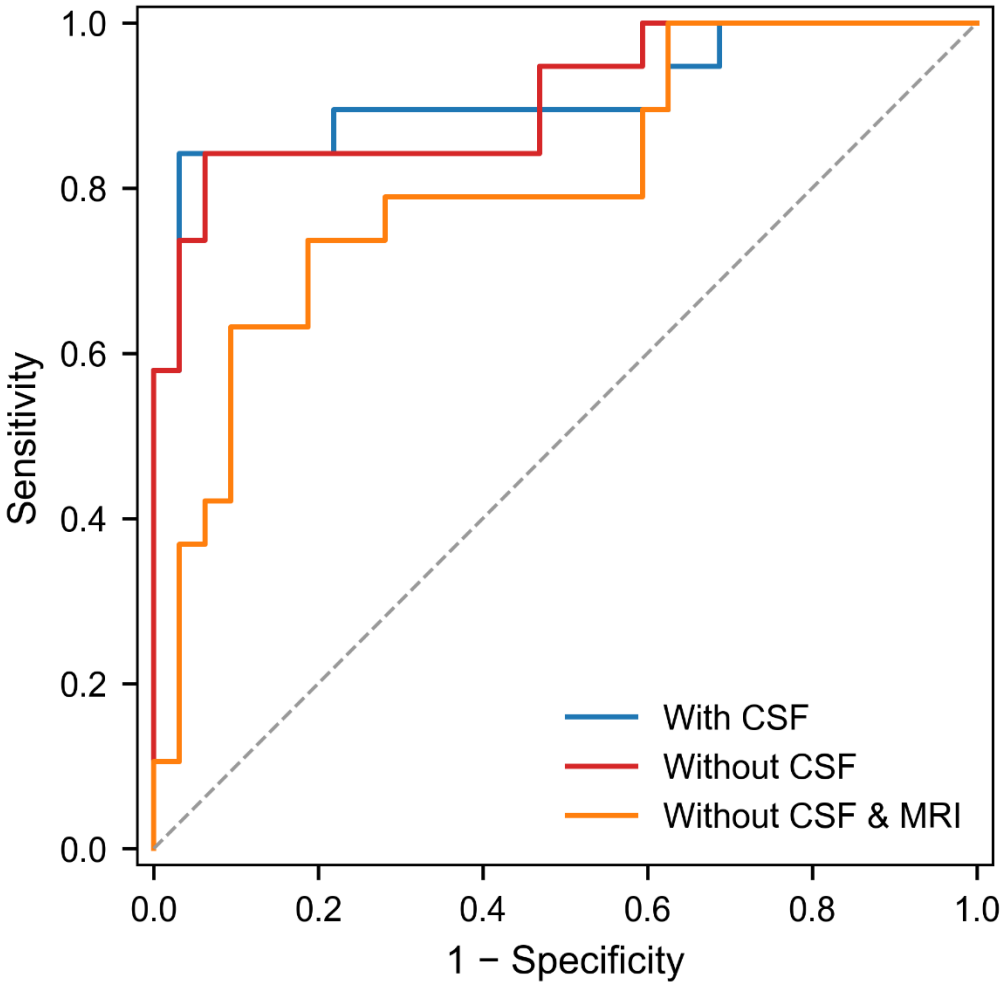
A. Discrimination (nested cross-validation, 5 outer folds × 20 repeats)

Feature set	Features, n	AUC, mean (SD)	2.5–97.5th percentile	AUC, repeat-averaged predictions
With CSF	29	0.894 (0.031)	0.842–0.943	0.911
Without CSF	25	0.902 (0.024)	0.854–0.937	0.908
Without CSF and MRI	23	0.769 (0.030)	0.715–0.810	0.806

B. Pairwise comparison (DeLong test on repeat-averaged out-of-fold probabilities)

Comparison (A vs B)	ΔAUC (A − B)	95% CI	p value
With CSF vs Without CSF	+0.003	−0.033 to +0.040	0.86
Without CSF vs Without CSF and MRI	+0.102	−0.017 to +0.221	0.094
With CSF vs Without CSF and MRI	+0.105	−0.023 to +0.234	0.108

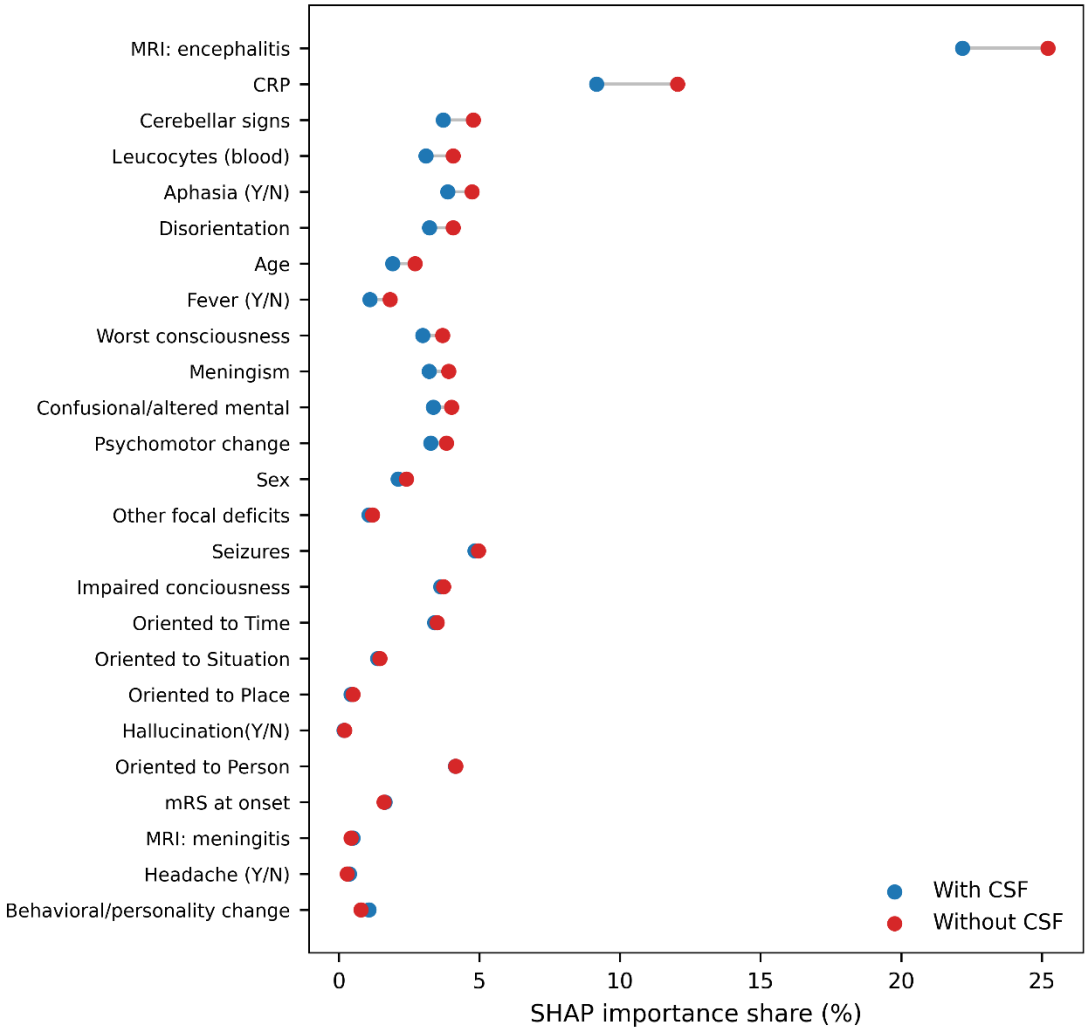
All settings were identical to the main analysis except for the feature set. In panel A, "AUC, mean (SD)" is the mean (SD) of the per-repeat pooled out-of-fold AUC, and "AUC, repeat-averaged predictions" is the AUC of the per-case out-of-fold probabilities averaged across repeats (the curves plotted in Fig. S2). In panel B, ΔAUC, its 95% CI, and the p value are from the DeLong test applied to the same repeat-averaged probabilities (an approximate paired comparison; a positive ΔAUC favours the first-named set); ΔAUC therefore equals the difference between the repeat-averaged AUCs in panel A, not the difference between the mean per-repeat AUCs. The "Without CSF and MRI" set is an exploratory sensitivity analysis, not a proposed clinical model. AUC, area under the ROC curve; CSF, cerebrospinal fluid; MRI, magnetic resonance imaging.



Supplementary Figure S2.

Receiver operating characteristic (ROC) curves for the three feature sets: with cerebrospinal fluid (CSF) features (29 features), without CSF features (25), and without CSF and MRI features (23). Curves are derived from out-of-fold predicted probabilities averaged across the repeats of the nested cross-validation (5 outer folds × 20 repeats); the dashed diagonal denotes chance-level performance. As in Fig. 1, AUC values are not shown on the plot; the AUCs of the plotted averaged curves were 0.911 with CSF, 0.908 without CSF, and 0.806 without CSF and MRI (Supplementary Table S2). The "without CSF and MRI" model is an exploratory sensitivity analysis assessing the contribution of MRI findings, not a proposed clinical model. CSF, cerebrospinal fluid; MRI,

magnetic resonance imaging.



Supplementary Figure S3.

Comparison of SHAP importance shares between the separately tuned models with and without CSF features. For the 25 non-CSF features common to both models, points show each feature's SHAP importance share (percentage of total mean absolute SHAP) in the model with CSF (blue) and without CSF (red), connected by a line. The two models showed different relative importance profiles, with higher shares for encephalitis-typical MRI findings and C-reactive protein in the CSF-free model. Feature names are abbreviated; full definitions are given in Table 1.