# SPIKE V2.0.1 · held-out results per event / artefact

Patients never used for training or tuning · scored 29 Oct 2026 · seizure rule re-scored 17 Nov 2026 (see †)

| | ROC AUC | | | False alarms per hour of EEG | | |
|---|---|---|---|---|---|---|
| | Trees | Deep learning | **Vote** (used in app) | Trees | Deep learning | **Vote** (used in app) |
| **Epileptiform events** | | | | | | |
| Seizure (V2.0 rule)<br>416 events · 278 h EEG | 0.88 | 0.89 | **0.90** | 7.9<br>84 % caught | 6.3<br>83 % caught | 6.0<br>83 % caught |
| **Seizure (V2.0.1 rule) †**<br>≥ 10 s, peak ≥ 0.625 · 416 events | same | same | same | – | – | **2.2**<br>76 % caught |
| GPD<br>58 events · 33 h EEG | 0.91 | **0.96** | **0.96** | 28.5<br>100 % caught | 13.4<br>100 % caught | 12.2<br>100 % caught |
| LPD<br>224 events · 33 h EEG | 0.80 | **0.86** | **0.86** | 27.9<br>83 % caught | 14.8<br>92 % caught | 16.6<br>88 % caught |
| Spike-wave<br>23 events · 33 h EEG | **0.81** | 0.79 | **0.81** | 128<br>74 % caught | 173<br>65 % caught | 156<br>70 % caught |
| **Artefacts** | | | | | | |
| Eye movement<br>2,045 events · 52 h EEG | 0.90 | 0.90 | **0.91** | 68.4<br>78 % caught | 91.1<br>92 % caught | 75.0<br>89 % caught |
| Muscle<br>1,371 events · 19 h EEG | **0.89** | 0.85 | **0.89** | 22.7<br>71 % caught | 28.8<br>78 % caught | 18.7<br>73 % caught |
| Electrode<br>787 events · 19 h EEG | 0.79 | 0.79 | **0.80** | 73.3<br>74 % caught | 92.7<br>84 % caught | 71.7<br>78 % caught |
| Chewing<br>59 events · 19 h EEG | **0.95** | 0.93 | **0.95** | 28.1<br>83 % caught | 4.8<br>51 % caught | 3.2<br>68 % caught |
| Shivering<br>1 holdout patient · hidden in app | 0.57 | 0.66 | **0.69** | – | – | – |
| Any artefact<br>2,885 events · 52 h EEG | 0.85 | 0.84 | **0.86** | 87.1<br>83 % caught | 91.9<br>91 % caught | 78.8<br>88 % caught |

† Designer leakage. The seizure holdout has been looked at twice: the V2.0.1 seizure rule was designed after the first holdout results (29 Oct 2026) had been seen, and the same holdout patients were then scored again (17 Nov 2026). Treat the seizure holdout numbers as optimistic. The rule cut false alarms from 7.6 to 2.2 per hour and seizures caught from 85 % to 76 % (95 % CI 56–91 %), both with the same scoring code; the V2.0 row above comes from the original evaluation code.

ROC AUC: how well each model separates event from non-event, per 2-s EEG window (1.00 = perfect, 0.50 = chance). Bold = best in the row. False alarms per hour: wrong detections per hour of complete holdout recordings, at the threshold chosen on development data (target: 70 % of events caught, spike-wave 40 %). Read it together with the % of holdout events caught underneath. Vote = the 50/50 average of the trees and the deep-learning model; epileptiform rows include artefact weighting, as in the app. Spike-wave: in the app a morphology check follows the vote (17–158 false alarms/h at 13–83 % caught, depending on the slider).