Received: 23 December 2022 | Revised: 30 June 2023 | Accepted: 5 July 2023

DOI: 10.1111/epi.17710

Epilepsia™

RESEARCH ARTICLE

Machine learning using multimodal clinical, electroencephalographic, and magnetic resonance imaging data can predict incident depression in adults with epilepsy: A pilot study

Guillermo Delgado-García[1,2] | Jordan D. T. Engbers[3] | Samuel Wiebe[1,2,4,5,6] | Pauline Mouches[7] | Kimberly Amador[7] | Nils D. Forkert[1,2,7] | James White[7,8,9] | Tolulope Sajobi[1,2,4,5] | Karl Martin Klein[1,2,4,10,11] | Colin B. Josephson[1,2,4,5,12] | on behalf of the Calgary Comprehensive Epilepsy Program Collaborators

[1]Department of Clinical Neurosciences, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

[2]Hotchkiss Brain Institute, University of Calgary, Calgary, Alberta, Canada

[3]Desid Labs, Calgary, Alberta, Canada

[4]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

[5]O'Brien Institute for Public Health, University of Calgary, Calgary, Alberta, Canada

[6]Clinical Research Unit, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

[7]Department of Radiology, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

[8]Libin Cardiovascular Institute, University of Calgary, Calgary, Alberta, Canada

[9]Department of Cardiac Sciences, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

[10]Department of Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

[11]Alberta Children's Hospital Research Institute, University of Calgary, Calgary, Alberta, Canada

[12]Centre for Health Informatics, University of Calgary, Calgary, Alberta, Canada

Correspondence
Colin B. Josephson, Cumming School of Medicine, Foothills Medical Centre, University of Calgary, 1403-29 St. NW, Calgary, AB T2N 2T9, Canada.
Email: cbjoseph@ucalgary.ca

Funding information
Epilepsy Canada; ERA PerMed, Grant/Award Number: ERAPERMED2018-134 RAISE-GENIC; Hotchkiss Brain Institute, University of Calgary

Abstract

Objective: This study was undertaken to develop a multimodal machine learning (ML) approach for predicting incident depression in adults with epilepsy.

Methods: We randomly selected 200 patients from the Calgary Comprehensive Epilepsy Program registry and linked their registry-based clinical data to their first-available clinical electroencephalogram (EEG) and magnetic resonance imaging (MRI) study. We excluded patients with a clinical or Neurological Disorders Depression Inventory for Epilepsy (NDDI-E)-based diagnosis of major depression at baseline. The NDDI-E was used to detect incident depression over a median of 2.4 years of follow-up (interquartile range [IQR] = 1.5–3.3 years). A ReliefF algorithm was applied to clinical as well as quantitative EEG and MRI parameters for feature selection. Six ML algorithms were trained and tested using

Guillermo Delgado-García and Jordan D. T. Engbers are co-first authors.

stratified threefold cross-validation. Multiple metrics were used to assess model performances.

Results: Of 200 patients, 150 had EEG and MRI data of sufficient quality for ML, of whom 59 were excluded due to prevalent depression. Therefore, 91 patients (41 women) were included, with a median age of 29 (IQR = 22–44) years. A total of 42 features were selected by ReliefF, none of which was a quantitative MRI or EEG variable. All models had a sensitivity > 80%, and five of six had an F1 score ≥ .72. A multilayer perceptron model had the highest F1 score (median = .74, IQR = .71–.78) and sensitivity (84.3%). Median area under the receiver operating characteristic curve and normalized Matthews correlation coefficient were .70 (IQR = .64–.78) and .57 (IQR = .50–.65), respectively.

Significance: Multimodal ML using baseline features can predict incident depression in this population. Our pilot models demonstrated high accuracy for depression prediction. However, overall performance and calibration can be improved. This model has promise for identifying those at risk for incident depression during follow-up, although efforts to refine it in larger populations along with external validation are required.

KEYWORDS

depression, EEG, epilepsy, machine learning, MRI, prediction

1 | INTRODUCTION

Up to 25%–30% of people with epilepsy are diagnosed with depression.[1] Furthermore, there is a bidirectional association between these two conditions.[2] It is estimated that the hazard of depression in people with epilepsy is twofold higher than the general population.[2] Due to a focus on seizures, the defining manifestation of epilepsy, often ancillary comorbidities such as depression go underdiagnosed in routine clinic settings.[1]

Therefore, it is crucial to anticipate and accurately identify those who will develop depression after a diagnosis of epilepsy to mitigate risk and initiate prompt treatment. Early diagnosis and treatment are associated with better mental health outcomes.[3] Attempts have been made to predict who will develop depression after temporal lobe epilepsy surgery.[4] However, there is currently no such tool for non-surgically treated patients seen at their first tertiary care clinic visit. This is critically important, because depression has been linked to worse seizure outcomes[2] and is frequently undertreated in people with epilepsy.[5] Depression has also been associated with lower quality of life (QoL)[6,7] and health state valuation,[8] and it may represent a more accurate predictor of both QoL and health state valuation than seizure frequency.[6–8]

Machine learning (ML) can identify complex patterns in high-dimensional data that are not directly evident to human observers and can be used to predict future events and outcomes.[9–11] Thus, a multisource ML approach could prove useful to predict which patients with epilepsy will develop incident depression. Such knowledge can provide the treating team with time to mitigate risk and initiate early treatment. Hence, this study aimed to use demographic, clinical, patient-reported outcome measure (PROM), electroencephalographic (EEG), and magnetic resonance imaging (MRI) data from epilepsy patients referred to their first clinic visit to build ML-based

KEY POINTS

- Incident depression is common following a diagnosis of epilepsy
- It is feasible to predict incident depression in adults with epilepsy by means of multimodal machine learning
- Our pilot models achieved high accuracy for incident depression prediction
- Efforts to refine performance and optimize calibration and external validation are required next steps
- Once validated, this model has promise for deployment in routine clinical practice to predict incident depression

multimodal prediction models to determine who will develop incident depression at follow-up visits.

2 | MATERIALS AND METHODS

2.1 | Data source, setting, and participants

Due to resource constraints, we used a convenience sample of 200 patients randomly selected from the Calgary Comprehensive Epilepsy Program (CEP) registry. The CEP is a level 4 epilepsy center in Western Canada. The CEP database comprises a prospective registry of adult (≥18 years of age) outpatient encounters covering a local catchment area of >1.3 million people. The CEP registry contains highly granular, routinely collected longitudinal demographic, clinical, PROM, and diagnostic data on every clinical encounter from almost 7000 people with epilepsy. Prospective data collection began in 2007, and data are acquired using standardized forms completed by both patients and staff epileptologists. Patients and epileptologists complete separate forms, with questions covering demographics, seizure and epilepsy characteristics, medical and psychiatric history, physical examination findings, results of diagnostic investigations, multidisciplinary consultations, use of antiseizure medications (ASMs), and surgical history. Epileptologists record epilepsy and seizure types based on all available data collected to that date. The epilepsy diagnosis is made by one of seven core staff epileptologists initially using the 2005 International League Against Epilepsy consensus definition[12] and then the updated 2014 criteria.[13] In this study, no surgical patients were included (e.g., those referred to Calgary from another tertiary program), to minimize heterogeneity and reflect practice for those first presenting for initial care as a first consult at a tertiary care center.

2.2 | Exposure and outcomes

Exposure was the diagnosis of epilepsy. The outcome was incident depression. Baseline (or prevalent) depression was defined as a clinician diagnosis of major depression or a Neurological Disorders Depression Inventory for Epilepsy (NDDI-E) score of >13[14] at the first clinic visit. In our center, the NDDI-E is completed by all patients at every clinical encounter, and therefore we were able to use it to identify both incident and prevalent depression. Thus, our study definition for incident depression was (1) no history of depression prior to or at the baseline CEP clinic visit as per the consultant physician's assessment, (2) an NDDI-E score of ≤13 at the baseline clinic visit, and (3) an NDDI-E score of >13[15] at any point during patient follow-up.

2.3 | Predictors

Demographic, clinical, and PROM data from the first clinic visit, driven by clinical acumen and literature reviews, were included to develop a model that could be applied at initial clinical assessment. Demographic variables included age, sex, marital status, employment status, and need for provincial and/or federal income assistance. Clinical variables included alcohol, tobacco, and recreational drug use, seizure history, epilepsy risk factors, seizure triggers, ASMs (including number, relative defined daily dose [ratio of the daily dose the patient is taking and the World Health Organization recommended daily dose], and presence of side effects), dietary therapy, medical and psychiatric comorbidities, abnormal general physical and neurological examination findings, seizure classification and frequency, and epilepsy type and potential localization. Clinical interpretation of MRI and EEG studies, as interpreted by consultant-level neuroradiologists and electroencephalographers, respectively, were also included. In the CEP, eight PROMs are collected at every clinic visit, including the NDDI-E, EuroQol 5-Dimension 5-Level instrument (EQ-5D-5L), Global Assessment of Disability of Epilepsy (GAD), Global Assessment of Severity of Epilepsy (GASE), Patient-Weighted Quality of Life in Epilepsy Inventory-10 (QOLIE-10-P), Liverpool Adverse Events Profile, seven-item Generalized Anxiety Disorder scale, and Treatment Satisfaction Questionnaire for Medication.[16]

We obtained the patient's first available MRI and EEG for signal analysis. The clinical EEG interpretations were obtained from the chart and are all reported by Canadian Society for Clinical Neurophysiology board-certified neurologists as part of routine care. The quantitative EEG data were extracted from these same recordings. All EEGs were performed using the International 10–20 system[17] and included at least 30 min of recording. With the 60-Hz notch filter applied, we manually isolated 10-s epochs of eyes-closed, relaxed wakefulness. Multiple epochs per patient were used (maximum of six per patient), replicating a previously published protocol for using quantitative EEG to predict seizure freedom.[18] EEG features were selected based on heuristics and theoretical frameworks. The quantitative EEG features included global explained variance of the EEG microstate maps, EEG microstate global field power peaks per second, Shannon entropy and entropy rate of the distribution of EEG microstates,[19] spectral entropy from electrode pairs, and power of beta and theta frequencies from electrode pairs. Electrode

pairs comprised prefrontal (Fp1/Fp2), frontal (F3/F7, F4/F8), temporal (T7/T8), central (C3/C4), parietal (P3/P4), and occipital (O1/O2) regions. Quantitative EEG features (Table S1) were computed using MNE-Python version 1.0.2[20] and the open-source Python package eeg_microstates.py.[19]

We obtained the closest available brain MRI datasets in DICOM format to the patient's first clinic visit and computed the regional brain volume (mm^3) and surface area (mm^2) of 27 brain regions of interest (Table S2) from T1-weighted images. The images were resampled to isotropic resolution (256×256×256) using FreeSurfer,[21] and then segmented into 27 regions of interest using QuickNAT,[22] a deep learning-based neuroanatomy segmentation tool. This brain parcellation method was selected because it results in a robust segmentation of a comprehensive number of brain areas compared to the subjects available and includes the most important deep gray matter structures that are known to be affected in people with depression.[23] The number of voxels present in each region was used to determine its volume, and the surface area was estimated using a run-length encoding-based algorithm.[24] These volume and surface area estimations were done using the Insight Toolkit.[25] Median apparent diffusion coefficient (ADC) values from diffusion-weighted imaging (DWI) MRI sequences were also extracted from the same regions. Therefore, the affine registration transformation between the T1-weighted and DWI datasets of each patient was estimated and used to transform the segmented regions of interest from the T1-weighted MRI space into DWI space. These transformed segmentations were then used as masks to compute the median ADC values in each region. Finally, the total brain volume of each patient was used as an additional predictor. This was estimated by using a total brain mask defined in the Montreal Neurological Institute brain template[26] space, which was propagated to each patient's T1-weighted MRI space using a nonlinear registration transformation. All image registration steps were performed using the Advanced Normalization Tools toolkit.[27]

2.4 | Statistical analysis and ML

A complete case analysis was used. Descriptive data are presented as median (interquartile range [IQR]) or count (%). The ReliefF feature selection algorithm[28] was applied on demographic, clinical, PROM, EEG, and MRI features. Briefly, a ReliefF algorithm cycles through m random training instances. With each cycle, the distance between a "target instance" (someone who develops incident depression) and all other instances is calculated. The target instance is then compared to its k nearest neighbors—one with the same class (developed incident depression; "nearest hit") and one with the opposite class (remained free of depression; "nearest miss"). Use of k nearest neighbors increases weight estimate reliability, especially in "noisy" datasets, as can be encountered among those describing routine clinical practice. During each step, the feature score vectors W are updated whereby if the value of that variable is different, then the weight increases by $1/m$ for the nearest miss and reduces by $1/m$ for the nearest hit. We then selected all features with weights of >0 for inclusion in the ML models.[29] Six ML classifiers were trained using the variables selected by ReliefF and tested using 100 iterations of stratified threefold cross-validation. Median and IQR for model metrics were reported following these iterations.

The following ML classifiers were used in this study: support vector machine using a polynomial kernel, multilayer perceptron, gradient boosting consensus, random forest, Gaussian Naive Bayes, and L2 penalized logistic regression. Discrimination, calibration, and model performance were evaluated using the F1 score, normalized Matthews correlation coefficient (MCC), area under the receiver operating characteristic curve (AUC), sensitivity, and specificity for each classifier. The F1 score is the harmonic mean of the positive predictive value and sensitivity, with a range of 0 to 1, where 0 means no correct classifications and 1 means perfect positive predictive value and sensitivity; the normalized MCC evaluates model performance (discrimination and calibration), with a range of 0 to +1, where 0 is no agreement between prediction and observation and +1 is perfect agreement (using the formula [MCC + 1]/2, which linearly projects the original range into a zero-to-one interval). These analyses were conducted using scikit-learn[30] in Python version 3.6.3 (Python Software Foundation).

2.5 | Ethics

Ethics approval for this study was obtained through both the University of Calgary's Conjoint Health Research Ethics Board and Alberta Health Services (REB18-0540_REN4). All CEP data are collected, managed, stored, and extracted using REDCap,[31] an electronic data capture tool hosted by the Clinical Research Unit at the University of Calgary.

3 | RESULTS

Of the 200 randomly selected patients, 150 (75%) had EEG signal (164/200 EEGs) and MRI data (173/200 MRIs) of sufficient quality for adequate feature extraction. Of

those, 59 patients were excluded due to prevalent depression. Therefore, 91 (61%) patients (41 [45%] female) were included in the analysis (Table 1). Their median age was 29 years (IQR = 22–44), and 18% were 1-year seizure-free at baseline. Focal epilepsy was diagnosed in 66 (72.5%) patients, and the rest had generalized epilepsy. These patients were first seen in clinic between 2014 and 2019, with a median of three follow-up visits (IQR = 2–4) over a median of 2.4 years (IQR = 1.5–3.3 years). The first available EEG was performed a median of 4 months prior to the clinic visit (IQR = 20 months prior to the clinic visit to 1 month after the clinic visit), and the first available MRI was performed a median of 9 months prior to the clinic visit (IQR = 2.9 years prior to the clinic visit to 1 month prior to clinic visit).

A total of 17 (19%) patients developed incident depression. Those who developed incident depression had an additional two clinic visits (median = 5 [IQR = 4–6] vs. 3 [IQR = 2–4], $p = .03$), although follow-up time was comparable (median = 2.8 years vs. 2.3 years), had more reported cannabis use (47% vs. 18%, $p = .014$), had a higher percentage reporting focal aware seizures (59% vs. 23%, $p = .006$), were less likely to have achieved 1-year seizure freedom prior to the index clinic visit (0% vs. 23%, $p = .020$), and had lower QoL (median QOLIE-10-P = 52 [IQR = 28–65] vs. 67 [IQR = 53–88], $p = .003$) and health state valuation (median EQ-5D-5L time trade-off = .87 [IQR = .82–.90] vs. .92 [IQR = .89–.94]).

3.1 | Feature selection

A total of 42 features had a ReliefF score of >0, suggesting predictive capacity, and were therefore included in ML models (Table S3). All quantitative MRI and EEG variables were filtered out by ReliefF algorithm, and thus were not included in the ML models. Included variables comprised demographics, comorbidities, substance use, epilepsy risk factors, epilepsy type (including structural epilepsy), seizure-related factors, EEG abnormalities (as described in the clinical report), social factors, and specific PROMs (Table S3). Among comorbidities, different psychiatric, neurodevelopmental, and medical conditions were identified, including the composite epilepsy-specific comorbidity index score.[32] Alcohol misuse, cigarette smoking, cannabis, and use of other substances were also selected as relevant features. Among seizure-related factors, seizure frequency for different seizure types, ASM change at first visit, baseline seizure freedom, use of enzyme-inducing ASM, polytherapy, and ASM side effects were identified. Reports of focal and generalized interictal epileptiform discharges as well as focal and generalized slowing on EEGs were also predictive. Relevant social factors identified by the ReliefF algorithm included employment status (including retirement) and the need for financial support through social services. Finally, baseline QOLIE-10-P subscores (driving, work, seizure worry, mental effect of ASMs, and physical effect of ASMs), GASE, and GAD[16] were also selected.

3.2 | Model performance

Overall, the performance of the models consistently favored a higher sensitivity at the expense of specificity (Figure 1). Our six ML-based models had a sensitivity of >80% (Table 2) for incident depression prediction. In addition, except for the Gaussian Naive Bayes algorithm, all ML models had an F1 score of >.71 (Table 2). The multilayer perceptron (Table 2) was the model with the highest F1 score (median = .74, IQR = .71–.78) and sensitivity (84.3%). Discrimination (AUC) and calibration/performance (MCC) of this model were moderate, at .70 (IQR = .64–.78) and .57 (IQR = .50–.65), respectively (Table 2).

4 | DISCUSSION

Our pilot study showed that it is feasible to link demographic, clinical, and PROM data to quantitative EEG and MRI data for predictive modeling to create multisource models in epilepsy. Likewise, our study shows that ML algorithms using available features from the first clinic visit were able to predict future incident depression in non-surgically treated patients with clinically relevant accuracy. This is likely in part because clinically informed models tend to outperform those solely dependent on data-driven methods.[33] These models demonstrate moderate to high discriminative capacity, which is ideal for a screening test. Here, the goal is to identify patients at high risk for depression, thus permitting targeted follow-up and management through confirmatory testing and treatment.

The discriminative capacity is not surprising, because the predictors identified by the ReliefF algorithm (Table S3) are clinically intuitive and consistent with prior reports. For instance, anxiety[34,35] and attention-deficit/hyperactivity disorder[36] are well-known risk factors for incident depression. Also, alcohol use disorders have been associated with an increased risk of subsequent depressive symptoms.[37] Some of the seizure- and ASM-related predictors identified, such as the need for polytherapy, are also reportedly associated with emergent depression in people with epilepsy.[35]

Interestingly, the quantitative MRI and EEG features were not selected by the ReliefF algorithm, a phenomenon that could be explained in two ways. First, clinical factors

TABLE 1 Baseline demographic, clinical, and patient-reported outcome measure variables of study cohort at their first visit to the Calgary CEP ($N = 91$).

Feature	Total, $N = 91$	No depression, $n = 74$	Incident depression, $n = 17$	p
Age, years, median (IQR)	29 (22–44)	30.5 (21.2–46)	24 (22–36)	.273
Sex, female, *n* (%)	41 (45)	30 (41)	11 (65)	.071
Total visits, median (IQR)	3 (2–4)	3 (2–4)	5 (4–6)	<.001
Follow-up, years, median (IQR)	2.4 (1.5–3.3)	2.3 (1.4–3.2)	2.8 (2.4–3.6)	.03
Time to EEG, months, median (IQR)[a]	-4 (−20 to 1)	−3 (−18 to 2)	−12 (−24 to −4)	.11
Time to MRI, months, median (IQR)[a]	−9 (−34 to −1)	−8 (−34 to −1)	−8 (−35 to −3)	.91
Unemployed, *n* (%)	27 (30)	22 (30)	5 (29)	.979
Level of education, *n* (%)				
College level or higher	49 (54)	40 (54)	9 (53)	.572
Current driver, *n* (%)	43 (47)	38 (51)	5 (29)	.102
Alcohol use, *n* (%)	49 (53.8)	43 (58.1)	6 (35.2)	.076
Tobacco use, *n* (%)	15 (16.4)	11 (14.8)	4 (23.5)	.294
Cannabis use, *n* (%)	21 (23)	13 (17.5)	8 (47)	.014
Other recreational drug use, *n* (%)	3 (3.2)	2 (2.7)	1 (5.8)	
Epilepsy risk factors, *n* (%)				
Family history	24 (26.3)	20 (27)	4 (23.5)	.516
Severe head trauma	9 (9.8)	6 (8.1)	3 (17.6)	.220
Developmental disorder	6 (6.5)	5 (6.7)	1 (5.8)	.689
Perinatal injury	5 (5.4)	2 (2.7)	3 (17.6)	.043
Febrile seizures	4 (4.3)	3 (4)	1 (5.8)	.569
Epilepsy-specific medical comorbidity index				
Median (IQR), range	0 (0–0), 0–2	0 (0–0), 0–2	0 (0–0), 0–2	.799
Hypertension, *n* (%)	4 (5)	4 (5)	0 (0)	.327
Arrhythmias, *n* (%)	1 (1)	1 (1)	0 (0)	.630
Renal disease, *n* (%)	1 (1)	1 (1)	0 (0)	.630
Paraplegia/hemiplegia, *n* (%)	1 (1)	0 (0)	1 (6)	.036
Dementia, *n* (%)	0 (0)	0 (0)	0 (0)	1.0
Solid nonmetastatic tumor, *n* (%)	0 (0)	0 (0)	0 (0)	1.0
Aspiration pneumonia, *n* (%)	0 (0)	0 (0)	0 (0)	1.0
Brain tumor, *n* (%)	0 (0)	0 (0)	0 (0)	1.0
Anoxic brain injury, *n* (%)	0 (0)	0 (0)	0 (0)	1.0
Psychiatric history, *n* (%)				
Anxiety	6 (6.5)	3 (4)	3 (17.6)	.076
ADHD	2 (2.1)	1 (1.3)	1 (5.8)	.340
Bipolar disorder	1 (1)	0 (0)	1 (6)	.187
PTSD	1 (1)	0 (0)	1 (6)	.187
Intellectual disability	1 (1)	0 (0)	1 (6)	.187
Psychosis	0 (0)	0 (0)	0 (0)	(0%)
ASD	0 (0)	0 (0)	0 (0)	1.0

TABLE 1 (Continued)

Feature	Total, $N=91$	No depression, $n=74$	Incident depression, $n=17$	p
Epilepsy type, *n* (%)				
Focal	66 (73)	51 (69)	15 (88.2)	.108
Lesional[b]	11 (16.6)	9 (17.6)	2 (13.3)	
Generalized	25 (27)	23 (31)	2 (12)	.964
Seizure type, *n* (%)				
Bilateral/generalized tonic–clonic	63 (69.2)	49 (66.2)	14 (82.3)	.157
Focal aware	27 (29.6)	17 (22.9)	10 (58.8)	.006
Focal impaired awareness	26 (28.5)	20 (27)	6 (35.2)	.556
Myoclonic	6 (6.5)	5 (6.7)	1 (5.8)	.689
Typical absence	3 (3.2)	3 (4)	0 (0)	.534
ASM use at first visit, *n* (%)	78 (86)	63 (85)	15 (88)	.546
ASM polytherapy, *n* (%)	12 (16)	10 (15.8)	2 (13.3)	.604
ASM defined daily dose, median (IQR)	.66 (.6–1.2)	.66 (.6–1.2)	.66 (.5–1.3)	.729
1-year seizure-free at first visit, *n* (%)	17 (18.6)	17 (22.9)	0 (0)	.020
Patient-reported outcome measures				
GASE, *n* (%)				
Not at all severe [1]	3 (3.2%)	1 (1.3%)	2 (11.7%)	.162
A little to extremely severe [2–7]	88 (96.7%)	73 (98.6%)	15 (88.2%)	
GAD, *n* (%)				.012
Not at all disabling [1]	2 (2.1%)	1 (1.3%)	1 (5.8%)	
A little to extremely disabling [2–7]	89 (97.8%)	73 (98.6%)	16 (94.1%)	
Weighted QOLIE-10-P, median [IQR]	64.8 (47.9–87.5)	67.4 (53.4–88.3)	52.3 (28–64.8)	.003
EQ-5D-5L, median (IQR)				
TTO	.91 (.87–.94)	.92 (.89–.94)	.87 (.82–.9)	<.001
NDDI-E, median (IQR)	8 (6–10.5)	8 (6–10)	11 (8–13)	.001

Abbreviations: ADHD, attention-deficit/hyperactivity disorder; ASD, autism spectrum disorder; ASM, antiseizure medication; CEP, Comprehensive Epilepsy Program; EEG, electroencephalography; EQ-5D-5L, EuroQol 5-Dimension 5-Level instrument; GAD, Global Assessment of Disability of Epilepsy; GASE, Global Assessment of Severity of Epilepsy; IQR, interquartile range; MRI, magnetic resonance imaging; NDDI-E, Neurological Disorders Depression Inventory for Epilepsy; PTSD, posttraumatic stress disorder; QOLIE-10-P, Patient-Weighted Quality of Life in Epilepsy Inventory-10; TTO, time trade-off.

[a]Negative numbers mean prior to the baseline first CEP clinic visit.

[b]Lesional epilepsy is based on the presence of a concordant lesion on MRI.

including psychosocial and seizure- and epilepsy-related factors, as well as medical comorbidities, are highly associated with depression,[35] and their inclusion may supersede other features with comparatively less robust predictive capacity. QoL and health state valuation are highly correlated with the severity of prevalent depression, typically more so than seizure frequency,[38] and thus may also exert strong predictive power for incident depression. Additionally, clinical features are highly complex, and presuming they ultimately arise from underlying neuroanatomical and neurophysiological states, may include manifold latent variables including components of the underlying quantitative EEG and MRI data. If so, this could render the latter less informative due to variance inflation and multicollinearity. Second, the quantitative EEG and MRI features we selected a priori for this study may fail to exert predictive capacity either because they represent a single "snapshot" in time, the dynamics of which could have changed just prior to developing depression, or these specific features, although useful for predicting seizure freedom, may not exist on a causal pathway for depression in epilepsy. Evidence from our study may support

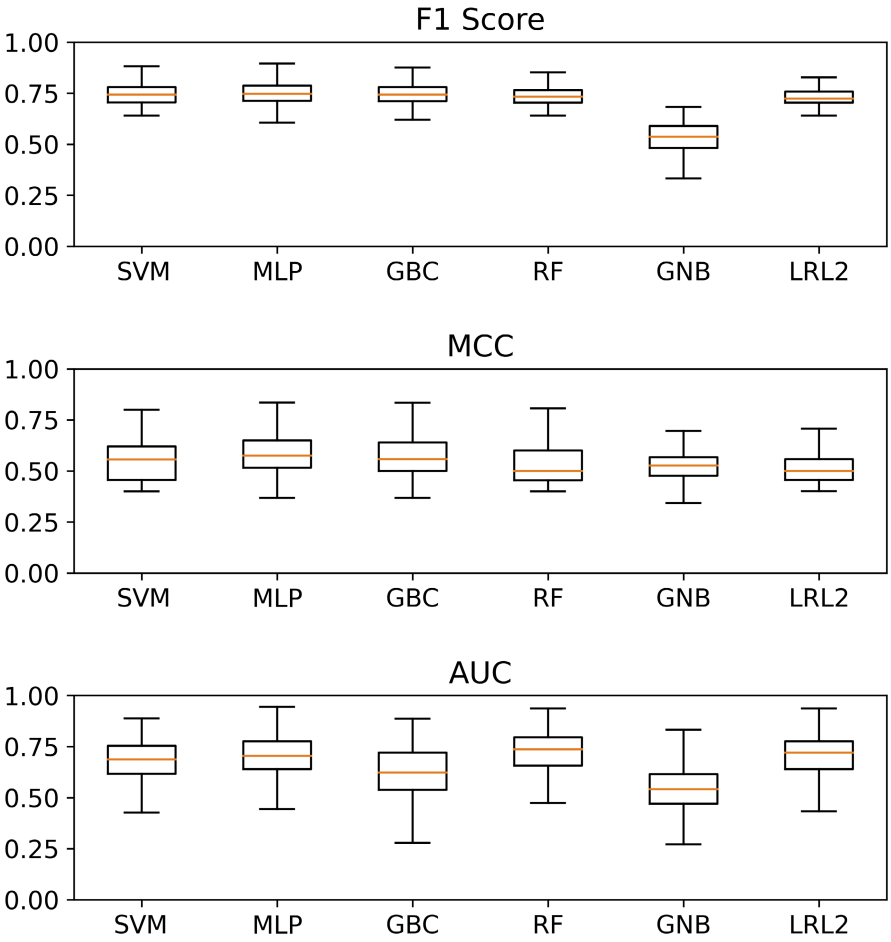

FIGURE 1 Boxplot of the median F1 score, normalized Matthews correlation coefficient (MCC), and area under the receiver operative curve (AUC) for each machine learning model that predicts incident depression in people with epilepsy following their first presentation to a tertiary care clinic. GBC, gradient boosting consensus; GNB, Gaussian Naive Bayes; LRL2, L2 penalized logistic regression; MLP, multilayer perceptron; RF, random forest; SVM, support vector machine.

TABLE 2 Performance of machine learning models for predicting incident depression following feature selection through a ReliefF algorithm.

Model	F1, median (IQR)	nMCC, median (IQR)	AUC, median (IQR)	Sen	Spe
SVM	.74 (.70–.78)	.56 (.45–.62)	.69 (.61–.75)	.830	.350
MLP	.74 (.71–.78)	.57 (.50–.65)	.70 (.64–.78)	.843	.348
GBC	.74 (.71–.78)	.56 (.50–.64)	.62 (.54–.72)	.835	.318
RF	.74 (.70–.77)	.50 (.46–.60)	.75 (.67–.80)	.821	.321
GNB	.53 (.48–.59)	.52 (.47–.57)	.54 (.46–.61)	.830	.200
LRL2	.72 (.70–.76)	.50 (.46–.56)	.72 (.64–.78)	.819	.300

Note: Included are measures of each model's accuracy (F1), performance (nMCC), and discrimination (AUC, Sen, Spe).

Abbreviations: AUC, area under the receiver operating characteristic curve; F1, F1 score; GBC, gradient boosting consensus; GNB, Gaussian Naive Bayes; IQR, interquartile range; LRL2, L2 penalized logistic regression; MLP, multilayer perceptron; nMCC, normalized Matthews correlation coefficient; RF, random forest; Sen, sensitivity; Spe, specificity; SVM, linear support vector machine.

the latter, as four different clinically reported EEG abnormalities (focal and generalized interictal epileptiform discharges as well as focal and generalized slowing) were selected as relevant predictors in our models. Exploration of other quantitative neurophysiological and anatomical features, as well as applying raw data to larger samples in deep learning approaches, could therefore improve model performance, even though clinical and psychosocial factors could be the primary drivers of depression in epilepsy.

We felt inclusion of EEG and MRI biomarkers was integral to investigate whether they are any more predictive of depression beyond the clinical variables. Although no quantitative EEG or MRI biomarker was selected by the final model, this work still has important clinical implications in that quantitative macro- and microstructural MRI biomarker data, as well as quantitative EEG data, may not be useful to acquire for this specific task. However, other biomarkers and complementary

imaging modalities, such as functional MRI, were not investigated and may still provide predictive information, which is beyond the scope of this work. However, this study adds value by informing these future studies.

Our study benefits from a well-defined population of people with epilepsy and granular information on epilepsy type, seizure types and frequency, treatment modalities, and comorbidities. We were able to use detailed demographic, clinical, and PROM data to accurately detect new onset depression, and well as comprehensively and consistently follow patients for up to 6.4 years (IQR = 1.5–3.3 years). The included population is representative of patients attending tertiary care epilepsy clinics in developed Western nations. We used well-described ML models only after automatic feature selection to reduce risks of overfitting. We also report multiple metrics—including F1 score, AUC, sensitivity, specificity, and normalized MCC—to permit readers the ability to fully evaluate each model.[39]

This study has some limitations that should be discussed. First, our population included only 91 patients, limited due to the need for convenience sampling related to restricted resources, the need to exclude people with prevalent depression (which is common in epilepsy due to their bidirectional relationship[2,40]), and data quality issues in real-world practice.[41,42] Increasing the sample size through acquisition of additional resources will likely result in greater precision and facilitate additional advanced analytic techniques such as deep learning. Deep learning will have the benefit of processing raw MRI and EEG signal data without the need for computing a priori features. Ultimately, exploration of other quantitative features, use of deep learning, and applying survival-based ML models could improve model performance. Imbalanced patient sample is another limitation of our analyses. In this context, rare outcomes are harder to predict, because there are fewer examples in the dataset. This imbalance is probably intrinsic and therefore resulted from naturally occurring frequencies,[40] as our sample was randomly selected from the CEP registry. Hence, it was important to report measures other than simple accuracy, such as the MCC,[39] to provide a transparent picture of model performance. We also relied on the NDDI-E to determine depression, which while having a high sensitivity (.81) and specificity (.90) for major depression, has a more modest positive predictive value (.62).[14] Although positive predictive value is dependent on prevalence, and therefore not immutable between populations, we cannot rule out that a plethora of false positive cases identified by the NDDI-E may have led to reduced specificity in our model.

ML algorithms using data derived from the first clinic visit are promising tools to predict incident depression in non-surgically treated patients with epilepsy. Such data can form the nidus of pipelines linking MRI and EEG studies for the purposes of prediction modeling. After application to feature and dimensionality reduction algorithms, such data can be used in ML classifiers to predict outcomes with increasing accuracy, thus galvanizing efforts to advance precision epilepsy. Specifically, our pilot models demonstrated a high accuracy for incident depression prediction, which is promising for potential use of the pipeline for other prediction and classification tasks in patients with epilepsy. These models are ideally situated as screening tools where high discriminatory capacity is desired. Despite this, future efforts are needed to refine and externally validate these algorithms before they are ready for clinical practice. Other quantitative MRI and EEG features, as well as the application of deep learning to larger samples, could add value to fully exploit these modalities for depression prediction in people with epilepsy. Finally, using the aforementioned approaches, we can also identify features strongly predictive of depression, thus permitting precision studies targeting early intervention in individual patients to improve outcomes. Based on these data, such endeavors are feasible and should be pursued with immediate effect.

AUTHOR CONTRIBUTIONS

Guillermo Delgado-García: Data acquisition; data analysis; interpretation of data; writing (original draft preparation; review and editing). **Jordan D. T. Engbers:** Analysis and interpretation of data; software and visualizations; writing (review and editing). **Samuel Wiebe, Karl Martin Klein:** Data acquisition; data analysis; interpretation of data; writing (review and editing). **Pauline Mouches, Kimberly Amador, Nils D. Forkert, James White, Tolulope Sajobi:** Analysis and interpretation of data; writing (review and editing). **Colin B. Josephson:** Conception of this project; data acquisition; data analysis; interpretation of data; writing (review and editing). **Calgary Comprehensive Epilepsy Program Collaborators:** Data acquisition. All authors approved the final version of the manuscript. The corresponding author is accountable for all aspects of this work.

ACKNOWLEDGMENTS

The Calgary Comprehensive Epilepsy Program collaborators include Drs. Paolo Federico, William Murphy, Neelan Pillay, Andrea Salmon, and Shaily Singh.

FUNDING INFORMATION

This work was supported by Epilepsy Canada, the Hotchkiss Brain Institute (University of Calgary) and the Canadian Institutes of Health Research (CIHR), under the frame of ERA PerMed. [Correction added on 05 August 2023, after first online publication: The Funding Information has been changed.]

CONFLICT OF INTEREST STATEMENT

G.D.-G. is a Level-1 National Researcher, Mexican System of Researchers, Mexican Council of Science and Technology. S.W. has received unrestricted educational grants from UCB Pharma, Eisai, and Sunovion for work unrelated to this project. J.W. has received research funding from Siemens Healthineers, Circle Cardiovascular, and Pfizer. K.M.K. reports personal fees from UCB Pharma, Novartis Pharma, Eisai, and GW Pharmaceuticals unrelated to this project, and grants from the federal state Hessen, Germany, through the LOEWE program, and from the Canadian Institutes of Health Research for work unrelated to this project. C.B.J. has received unrestricted educational grants from UCB Pharma and Eisai for work unrelated to this project. None of the other authors has any conflict of interest to disclose. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

ORCID

Guillermo Delgado-García https://orcid.org/0000-0002-3123-5879
Samuel Wiebe https://orcid.org/0000-0002-1061-9099
Tolulope Sajobi https://orcid.org/0000-0002-5696-5552
Karl Martin Klein https://orcid.org/0000-0002-6654-1665
Colin B. Josephson https://orcid.org/0000-0001-7052-1651

REFERENCES

1. Mula M, Kanner AM, Jetté N, Sander JW. Psychiatric comorbidities in people with epilepsy. Neurol Clin Pract. 2021 Apr;11(2):e112–20. https://doi.org/10.1212/CPJ.0000000000000874
2. Josephson CB, Lowerison M, Vallerand I, Sajobi TT, Patten S, Jette N, et al. Association of depression and treated depression with epilepsy and seizure outcomes: a multicohort analysis. JAMA Neurol. 2017 May 1;74(5):533–9. https://doi.org/10.1001/jamaneurol.2016.5042
3. Ghio L, Gotelli S, Marcenaro M, Amore M, Natta W. Duration of untreated illness and outcomes in unipolar depression: a systematic review and meta-analysis. J Affect Disord. 2014 Jan;152-154:45–51. https://doi.org/10.1016/j.jad.2013.10.002
4. Doherty C, Nowacki AS, Pat McAndrews M, McDonald CR, Reyes A, Kim MS, et al. Predicting mood decline following temporal lobe epilepsy surgery in adults. Epilepsia. 2021 Feb;62(2):450–9. https://doi.org/10.1111/epi.16800
5. Ajinkya S, Fox J, Lekoubou A. Trends in prevalence and treatment of depressive symptoms in adult patients with epilepsy in the United States. Epilepsy Behav. 2020 Apr;105:106973. https://doi.org/10.1016/j.yebeh.2020.106973
6. Boylan LS, Flint LA, Labovitz DL, Jackson SC, Starner K, Devinsky O. Depression but not seizure frequency predicts quality of life in treatment-resistant epilepsy. Neurology. 2004 Jan;62(2):258–61. https://doi.org/10.1212/01.wnl.0000103282.62353.85
7. Johnstone B, Malpas CB, Velakoulis D, Kwan P, O'Brien TJ. Psychiatric symptoms are the strongest predictors of quality of life in patients with drug-resistant epilepsy or psychogenic nonepileptic seizures. Epilepsy Behav. 2021 Apr;117:107861. https://doi.org/10.1016/j.yebeh.2021.107861
8. Josephson CB, Engbers JDT, Wang M, Perera K, Roach P, Sajobi TT, et al. Calgary comprehensive epilepsy program collaborators. Psychosocial profiles and their predictors in epilepsy using patient-reported outcomes and machine learning. Epilepsia. 2020 Jun;61(6):1201–10. https://doi.org/10.1111/epi.16526
9. MacEachern SJ, Forkert ND. Machine learning for precision medicine. Genome. 2021 Apr;64(4):416–25. https://doi.org/10.1139/gen-2020-0131
10. Lhatoo SD, Bernasconi N, Blumcke I, Braun K, Buchhalter J, Denaxas S, et al. Big data in epilepsy: clinical and research considerations. Report from the epilepsy big data task force of the international league against epilepsy. Epilepsia. 2020 Sep;61(9):1869–83. https://doi.org/10.1111/epi.16633
11. Josephson CB, Wiebe S. Precision medicine: academic dreaming or clinical reality? Epilepsia. 2021 Mar;62(Suppl 2):S78–S89. https://doi.org/10.1111/epi.16739
12. Fisher RS, van Emde BW, Blume W, Elger C, Genton P, Lee P, et al. Epileptic seizures and epilepsy: definitions proposed by the international league against epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). Epilepsia. 2005 Apr;46(4):470–2. https://doi.org/10.1111/j.0013-9580.2005.66104.x
13. Fisher RS, Acevedo C, Arzimanoglou A, Bogacz A, Cross JH, Elger CE, et al. ILAE official report: a practical clinical definition of epilepsy. Epilepsia. 2014 Apr;55(4):475–82. https://doi.org/10.1111/epi.12550
14. Gilliam FG, Barry JJ, Hermann BP, Meador KJ, Vahle V, Kanner AM. Rapid detection of major depression in epilepsy: a multicentre study. Lancet Neurol. 2006 May;5(5):399–405. https://doi.org/10.1016/S1474-4422(06)70415-X
15. Gill SJ, Lukmanji S, Fiest KM, Patten SB, Wiebe S, Jetté N. Depression screening tools in persons with epilepsy: a systematic review of validated tools. Epilepsia. 2017 May;58(5):695–705. https://doi.org/10.1111/epi.13651
16. Delgado-García G, Wiebe S, Josephson CB. The use of patient-reported measures in epilepsy care: the Calgary comprehensive epilepsy program experience. J Patient Rep Outcomes. 2021 Oct 12;5(Suppl 2):83. https://doi.org/10.1186/s41687-021-00356-4
17. Klem GH, Lüders HO, Jasper HH, Elger C. The ten-twenty electrode system of the international federation. The International Federation of Clinical Neurophysiology. Electroencephalogr Clin Neurophysiol Suppl. 1999;52:3–6.
18. Varatharajah Y, Joseph B, Brinkmann B, Morita-Sherman M, Fitzgerald Z, Vegh D, et al. Quantitative analysis of visually reviewed normal scalp EEG predicts seizure freedom following anterior temporal lobectomy. Epilepsia. 2022 Jul;63(7):1630–42. https://doi.org/10.1111/epi.17257
19. von Wegner F, Laufs H. Information-theoretical analysis of EEG microstate sequences in python. Front Neuroinform. 2018 Jun 1;12:30. https://doi.org/10.3389/fninf.2018.00030
20. Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, et al. MEG and EEG data analysis with

MNE-python. Front Neurosci. 2013 Dec 26;7:267. https://doi.org/10.3389/fnins.2013.00267
21. Fischl B. FreeSurfer. Neuroimage. 2012;62(2):774–81.
22. Guha Roy A, Conjeti S, Navab N, Wachinger C. Alzheimer's disease neuroimaging initiative. QuickNAT: a fully convolutional network for quick and accurate segmentation of neuroanatomy. Neuroimage. 2019 Feb 1;186:713–27. https://doi.org/10.1016/j.neuroimage.2018.11.042
23. Tae WS, Kim SS, Lee KU, Nam EC, Kim KW. Validation of hippocampal volumes measured using a manual method and two automated methods (FreeSurfer and IBASPM) in chronic major depressive disorder. Neuroradiology. 2008 Jul;50(7):569–81. https://doi.org/10.1007/s00234-008-0383-9
24. Lehmann G, Legland D. Efficient N-dimensional surface estimation using Crofton formula and run-length encoding. Insight J. 2012;852. https://doi.org/10.54294/wdu86d
25. Ibanez L, Schroeder W. The ITK software guide. Clifton Park, New York, USA: Kitware, Inc; 2003. http://www.itk.org/ItkSoftwareGuide.pdf
26. Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, et al. A probabilistic atlas and reference system for the human brain: international consortium for brain mapping (ICBM). Philos Trans R Soc Lond B Biol Sci. 2001 Aug 29;356(1412):1293–1322. https://doi.org/10.1098/rstb.2001.0915
27. Avants BB, Tustison N, Song G. Advanced normalization tools (ANTs). Insight J. 2009;2(365):1–25.
28. Robnik-Šikonja M, Kononenko I. In: Fisher DH editorAn adaptation of relief for attribute estimation in regression. Machine Learning: Proceedings of the Fourteenth International Conference (ICML97). Morgan Kaufmann Publishers Inc; 1997. p. 296–304.
29. Urbanowicz RJ, Meeker M, La Cava W, Olson RS, Moore JH. Relief-based feature selection: introduction and review. J Biomed Inform. 2018 Sep;85:189–203.
30. Pedregosa F, Varoquaux G, Gramfort A. Scikit-learn: machine learning in python. J Mach Learn Res. 2011;12(85):2825–30.
31. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009 Apr;42(2):377–81. https://doi.org/10.1016/j.jbi.2008.08.010
32. St Germaine-Smith C, Liu M, Quan H, Wiebe S, Jette N. Development of an epilepsy-specific risk adjustment comorbidity index. Epilepsia. 2011 Dec;52(12):2161–7. https://doi.org/10.1111/j.1528-1167.2011.03292.x
33. Steyerberg EW, Uno H, Ioannidis JPA, van Calster B, Collaborators. Poor performance of clinical prediction models: the harm of commonly applied methods. J Clin Epidemiol. 2018 Jun;98:133–43.
34. Jacobson NC, Newman MG. Anxiety and depression as bidirectional risk factors for one another: a meta-analysis of longitudinal studies. Psychol Bull. 2017 Nov;143(11):1155–1200. https://doi.org/10.1037/bul0000111
35. Yang Y, Yang M, Shi Q, Wang T, Jiang M. Risk factors for depression in patients with epilepsy: a meta-analysis. Epilepsy Behav. 2020 May;106:107030. https://doi.org/10.1016/j.yebeh.2020.107030
36. Riglin L, Leppert B, Dardani C, Thapar AK, Rice F, O'Donovan MC, et al. ADHD and depression: investigating a causal explanation. Psychol Med. 2021 Aug;51(11):1890–7. https://doi.org/10.1017/S0033291720000665
37. Li J, Wang H, Li M, Shen Q, Li X, Zhang Y, et al. Effect of alcohol use disorders and alcohol intake on the risk of subsequent depressive symptoms: a systematic review and meta-analysis of cohort studies. Addiction. 2020 Jul;115(7):1224–43. https://doi.org/10.1111/add.14935
38. Kanner AM. Management of psychiatric and neurological comorbidities in epilepsy. Nat Rev Neurol. 2016 Feb;12(2):106–16. https://doi.org/10.1038/nrneurol.2015.243
39. 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(1):6. https://doi.org/10.1186/s12864-019-6413-7
40. Rashid H, Upadhyay AD, Pandey RM, Katyal J. Point prevalence of depression in persons with active epilepsy and impact of methodological moderators: a systematic review and meta-analysis. Epilepsy Behav. 2021 Dec;125:108394. https://doi.org/10.1016/j.yebeh.2021.108394
41. Jiang X, Bian GB, Tian Z. Removal of artifacts from EEG signals: a review. Sensors (Basel). 2019 Feb 26;19(5):987. https://doi.org/10.3390/s19050987
42. Rados M, Mouthaan B, Barsi P, Carmichael D, Heckemann RA, Kelemen A, et al. Diagnostic value of MRI in the presurgical evaluation of patients with epilepsy: influence of field strength and sequence selection: a systematic review and meta-analysis from the E-PILEPSY consortium. Epileptic Disord. 2022 Apr 1;24(2):323–42.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Delgado-García G, Engbers JDT, Wiebe S, Mouches P, Amador K, Forkert ND, et al. Machine learning using multimodal clinical, electroencephalographic, and magnetic resonance imaging data can predict incident depression in adults with epilepsy: A pilot study. Epilepsia. 2023;64:2781–2791. https://doi.org/10.1111/epi.17710