## Appendix A

### 1. Data preprocessing appendix

**Table A.1.** Text features designed to build the structured text dataset for modeling.

|  |  |
| --- | --- |
| **Features type** | **Key words** |
| Bags of words for negative diagnosis of epilepsy {key word(s)} | {'not', 'evid', 'diagnosi', 'epilepsi'}, {'not', 'recommend', 'antiepilept', 'medic'}, {'defer', 'anti', 'seizur'}, {'defer', 'anti', 'epilept'}, {'referr', 'gener', 'neurolog'}, {'not', 'requir', 'follow', 'up'}, {'not', 'requir', 'followup'}, {'cannot', 'event', 'epilept'}, {'pnes'}, {'no', 'seizur', 'event'}, {'unlik', 'seizur'}, {'function', 'neurolog', 'disord'}, {'migrain'}, {'anxieti'}, {'syncop'}, {'convers', 'disord'}, {'psychogen'}, {'not', 'have', 'seizur', 'risk', 'factor'}, {'sleep', 'disord'}, {'sleep', 'apnea'}, {'not', 'recommend', 'test'}, {'low', 'suspicion', 'seizur'}, {'physiolog', 'tremor'}, {"''", 'seizur'}, {'function', 'neurolog'}, {'vasovag'}, {'defer', 'primary', 'care', 'physician'}, {'not', 'meet', 'definit', 'epilepsi'}, {'not', 'support', 'diagnosi', 'epilepsi'}, {'amnesia'}, {'provok', 'seizur'}, {'dispress'}, {'shiver'}, {'cardiac', 'arrest'}, {'no', 'anti', 'seizur', 'medic'}, {'neuropathi'}, {'neuropath'}, {'meningioma'}, {'hold', 'off', 'start', 'anti', 'epilept'}, {'diabet'}, {'neurosarcoidosi'}, {'sdh'}, {'post', 'oper'}, {'traumat', 'hemorrhag'}, {'low', 'concern', 'seizur'}, {'no', 'concern', 'seizur'}, {'not', 'convinc', 'seizur'}, {'not', 'need', 'follow', 'epilepsi'}, {'not', 'need', 'followup'}, {'not', 'start', 'antiepilept', 'medic'}, {'not', 'start', 'antiseizur', 'medic'}, {'unlik', 'epilepsi'}, {'trauma'}, {'traumat'}, {'hematoma'}, {'brain', 'abscess'}, {'hold', 'off', 'medic'}, {'postop'}, {'singl', 'seizur'}, {'singl', 'sz'}, {'function', 'event'}, {'follow', 'up', 'as', 'need'}, {'followup', 'as', 'need'}, {'referr', 'psychiatri'}, {'defer', 'medic'}, {'acut', 'symptomat', 'seizur'}, {'symptomat', 'seizur'}, {'first', 'time', 'seizur'}, {'one', 'lifetim', 'seizur'}, {'no', 'evid', 'seizur'}, {'not', 'meet', 'epilepsi'}, {'not', 'need', 'medic'}, {'jacobsen', 'syndrom'}, {'excess', 'alcohol'}, {'normal', 'neurolog', 'exam'}, {'normal', 'mri'}, {'normal', 'eeg'}, {'no', 'epilepsi', 'risk'}, {'no', 'epilepsi', 'risk', 'factor'}, {'no', 'epileptiform', 'abnorm'}, {'psychiatr'}, {'fentanyl'}, {'bipolar'}, {'not', 'have', 'epilepsi'}, {'no', 'bite'}, {'no', 'incontin'}, {'low', 'seizur', 'threshold'}, {'lower', 'seizur', 'threshold'}, {'no', 'antiseizur', 'medic'}, {'not', 'had', 'seizur'}, {'nonepilept'}, {'chemo'}, {'chemotherapi'}, {'no', 'epileptogen', 'abnorm'}, {'numb'}, {'surgeri'}, {'discharg', 'epilepsi', 'clinic'}, {'nonepileptiform'}, {'non', 'epileptiform'}, {'not', 'epilept'}, {'dementia'}, {'not', 'think', 'epilepsi'}, {'no', 'diagnosi', 'epilepsi'}, {'tingl'}, {'not', 'epileptiform', 'activ'}, {'no', 'seizur'}, {'withdraw', 'seizur'}, {'dizzi'}, {'maintain', 'conscious'}, {'no', 'electrograph', 'seizur'}, {'wean', 'off'}, {'taper'}, {'resect'}, {'second', 'opinion'}, {'definit', 'diagnosi', 'epilepsi'}, {'pseudoseizur'}, {'cardiolog'}, {'against', 'seizur'}, {'against', 'epilepsi'}, {'ptsd'}, {'psychogen', 'nonepilept', 'seizur'}, {'presyncop'}, {'hypoglycemia'}, {'doubt', 'seizur'}, {'not', 'carri', 'diagnosi', 'epilepsi'}, {'acut', 'seizur'}, {'deni', 'seizur'}, {'provok', 'spell'}, {'non', 'epilept', 'spell'}, {'nonepilept', 'spell'}, {'insomnia'}, {'migrain', 'aura'}, {'no', 'clinic', 'seizur'}, {'not', 'criteria', 'epilepsi'}. |
| Bags of words for positive diagnosis of epilepsy {key word(s)} | {'both', 'epilepsi', 'pnes'}, {'mix', 'disord'}, {'ictal'}, {'aura'}, {'convuls'}, {'breakthrough', 'seizur'}, {'focal'}, {'idiopath', 'general', 'epilepsi'}, {'histori', 'seizur'}, {'hx', 'seizur'}, {'complex', 'seizur'}, {'partial', 'seizur'}, {'myoclon'}, {'general', 'seizur'}, {'continu', 'on'}, {'drive', 'month'}, {'drive', 'seizur'}, {'deja', 'vu'}, {'seizurefre'}, {'szfree'}, {'seizur', 'free'}, {'sz', 'free'}, {'frontal', 'lobe'}, {'nocturn'}, {'febril'}, {'perinat', 'complic'}, {'seizur','control'}, {'monotherapi'}, {'absenc', 'seizur'}, {'dejavu'}, {'postict', 'confus'}, {'tonniclon'}, {'tonic', 'clonic'}, {'sudden', 'unexpect', 'death'}, {'sudep'}, {'facial', 'droop'}, {'intract', 'epilepsi'}, {'daili', 'seizur'}, {'decreas', 'seizur'}, {'devic'}, {'surgic', 'intervent'}, {'reprogram'}, {'abnorm', 'eeg'}, {'with', 'epilepsi'}, {'juvenil', 'epilespi'}, {'myoclonus'}, {'recurr', 'sz'}, {'recurr', 'seizur'}, {'noncompli'}, {'seizur', 'stabl'}, {'disloc', 'shoulder'}, {'narcolepsi'}, 'sleep clinic': {'sleep', 'clinic'}. |
| Anti-seizure medications | 'acetazolamid', 'acth', 'acthar', 'brivaracetam', 'briviact', 'cannabidiol', 'epidiolex', 'carbamazepin', 'cbz', 'epitol', 'tegretol', 'equetro', 'teril', 'carbatrol', 'tegretol', 'epitol', 'cenobam', 'xcopri', 'clobazam', 'frisium', 'onfi', 'sympazan', 'clonazepam', 'epitril', 'klonopin', 'rivotril', 'clorazep', 'tranxen', 'xene', 'diazepam', 'valium', 'diamox', 'diastat', 'divalproex', 'depakot', 'eslicarbazepin', 'aptiom', 'ethosuximid', 'zarontin', 'ethotoin', 'ezogabin', 'potiga', 'felbam', 'felbatol', 'gabapentin', 'neurontin', 'gralis', 'horiz', 'lacosamid', 'vimpat', 'lamotrigin', 'lamict', 'levetiracetam', 'ltg', 'ige', 'tpm', 'oxc', 'lev', 'keppra', 'roweepra', 'spritam', 'elepsia', 'lorazepam', 'ativan', 'methsuximid', 'methosuximid', 'celontin', 'oxcarbazepin', 'trilept', 'oxtellar xr', 'perampanel', 'fycompa', 'phenobarbit', 'luminol', 'lumin', 'phenytoin', 'epanutin', 'dilantin', 'phenytek', 'pregabalin', 'lyrica', 'primidon', 'mysolin', 'rufinamid', 'banzel', 'inovelon', 'percocet', 'stiripentol', 'diacomit', 'tiagabin', 'gabitril', 'topiram', 'topamax', 'topiram', 'qudexi', 'trokendi', 'valproat', 'valproic', 'wellbutrin', 'convulex', 'depacon', 'depaken', 'orfiril', 'valpor', 'valprosid', 'depakot', 'vigabatrin', 'sabril', 'vigadron', 'zonisamid', 'zonegran', 'xanax'. |

**Table A.2.** Final set of features included for modeling.

|  |  |
| --- | --- |
| **Feature type** | **Feature** |
| Text {bag of words key word(s)} | {'not', 'evid', 'diagnosi', 'epilepsi'}, {'not', 'recommend', 'antiepilept', 'medic'}, {'defer', 'anti', 'seizur'}, {{'not', 'requir', 'follow', 'up'}; {'not', 'requir', 'followup'}}, {'pnes'}, {'no', 'seizur', 'event'}, {'unlik', 'seizur'}, {'function', 'neurolog', 'disord'}, {'migrain'}, {'anxieti'}, {'syncop'}, {'convers', 'disord'}, {'psychogen'}, {'not', 'have', 'seizur', 'risk', 'factor'}, {'sleep', 'disord'}, {'sleep', 'apnea'}, {'not', 'recommend', 'test'}, {'low', 'suspicion', 'seizur'}, {'physiolog', 'tremor'}, {'function', 'neurolog'}, {'vasovag'}, {'not', 'meet', 'definit', 'epilepsi'}, {'not', 'support', 'diagnosi', 'epilepsi'}, {'amnesia'}, {'provok', 'seizur'}, {'shiver'}, {'cardiac', 'arrest'}, {'no', 'anti', 'seizur', 'medic'}, {'neuropathi'}, {'neuropath'}, {'meningioma'}, {'hold', 'off', 'start', 'anti', 'epilept'}, {'diabet'}, {'neurosarcoidosi'}, {'sdh'}, {'post', 'oper'}, {'traumat', 'hemorrhag'}, {'low', 'concern', 'seizur'}, {'no', 'concern', 'seizur'}, {'not', 'convinc', 'seizur'}, {'not', 'need', 'follow', 'epilepsi'}, {'not', 'need', 'followup'}, {'not', 'start', 'antiepilept', 'medic'}, {'not', 'start', 'antiseizur', 'medic'}, {'unlik', 'epilepsi'}, {'trauma'}, {'traumat'}, {'hematoma'}, {'brain', 'abscess'}, {'hold', 'off', 'medic'}, {'postop'}, {{'singl', 'seizur'}; {'singl', 'sz'}}, {'function', 'event'}, {{'follow', 'up', 'as', 'need'}; {'followup', 'as', 'need'}}, {'referr', 'psychiatri'}, {'defer', 'medic'}, {'acut', 'symptomat', 'seizur'}, {'symptomat', 'seizur'}, {'first', 'time', 'seizur'}, {'one', 'lifetim', 'seizur'}, {'no', 'evid', 'seizur'}, {'not', 'meet', 'epilepsi'}, {'not', 'need', 'medic'}, {'jacobsen', 'syndrom'}, {'excess', 'alcohol'}, {'normal', 'neurolog', 'exam'}, {'normal', 'mri'}, {'normal', 'eeg'}, {'no', 'epilepsi', 'risk'}, {'no', 'epilepsi', 'risk', 'factor'}, {{'no', 'epileptiform', 'abnorm'}; {'no', 'epileptogen', 'abnorm'}}, {'psychiatr'}, {'fentanyl'}, {'bipolar'}, {'not', 'have', 'epilepsi'}, {'no', 'bite'}, {'no', 'incontin'}, {{'low', 'seizur', 'threshold'}; {'lower', 'seizur', 'threshold'}}, {'no', 'antiseizur', 'medic'}, {'not', 'had', 'seizur'}, {'nonepilept'}, {{'chemo'}; {'chemotherapi'}}, {'numb'}, {'surgeri'}, {'discharg', 'epilepsi', 'clinic'}, {{'nonepileptiform'}; {'non', 'epileptiform'}}, {'not', 'epilept'}, {'dementia'}, {'not', 'think', 'epilepsi'}, {'no', 'diagnosi', 'epilepsi'}, {'tingl'}, {'not', 'epileptiform', 'activ'}, {'no', 'seizur'}, {'withdraw', 'seizur'}, {'dizzi'}, {'maintain', 'conscious'}, {'no', 'electrograph', 'seizur'}, {'wean', 'off'}, {'taper'}, {'resect'}, {'second', 'opinion'}, {'definit', 'diagnosi', 'epilepsi'}, {'pseudoseizur'}, {'cardiolog'}, {'against', 'seizur'}, {'against', 'epilepsi'}, {'ptsd'}, {'psychogen', 'nonepilept', 'seizur'}, {'presyncop'}, {'hypoglycemia'}, {'doubt', 'seizur'}, {'not', 'carri', 'diagnosi', 'epilepsi'}, {'acut', 'seizur'}, {'deni', 'seizur'}, {'provok', 'spell'}, {{'non', 'epilept', 'spell'}; {'nonepilept', 'spell'}}, {'insomnia'}, {'migrain', 'aura'}, {'no', 'clinic', 'seizur'}, {'not', 'criteria', 'epilepsi'}, {'both', 'epilepsi', 'pnes'}, {'mix', 'disord'}, {'ictal'}, {'aura'}, {'convuls'}, {'breakthrough', 'seizur'}, {'focal'}, {'idiopath', 'general', 'epilepsi'}, {{'histori', 'seizur'}; {'hx', 'seizur'}}, {'complex', 'seizur'}, {'partial', 'seizur'}, {'myoclon'}, {'general', 'seizur'}, {'continu', 'on'}, {'drive', 'month'}, {'drive', 'seizur'}, {{'dejavu'}; {'deja', 'vu'}}, {{'seizurefree'}; {'szfree'}; {'seizur', 'free'}; {'sz', 'free'}}, {'frontal', 'lobe'}, {'nocturn'}, {'febril'}, {'perinat', 'complic'}, {'seizur','control'}, {'monotherapi'}, {'absenc', 'seizur'}, {'postict', 'confus'}, {{'tonniclon'}; {'tonic', 'clonic'}}, {'sudden', 'unexpect', 'death'}, {'sudep'}, {'facial', 'droop'}, {'intract', 'epilepsi'}, {'daili', 'seizur'}, {'decreas', 'seizur'}, {'devic'}, {'surgic', 'intervent'}, {'reprogram'}, {'abnorm', 'eeg'}, {'with', 'epilepsi'}, {'myoclonus'}, {{'recurr', 'sz'}; {'recurr', 'seizur'}}, {'noncompli'}, {'seizur', 'stabl'}, {'disloc', 'shoulder'}, {'narcolepsi'}, {'sleep', 'clinic'}, {'acetazolamid'}, {'acth'}, {'acthar'}, {'brivaracetam'}, {'briviact'}, {'cannabidiol'}, {'epidiolex’}, {{'carbamazepin'}; {'cbz'}}, {'epitol'}, {'tegretol'}, {'carbatrol'}, {'cenobam'}, {'xcopri'}, {'clobazam'}, {'frisium'}, {'onfi'}, {'clonazepam'}, {'klonopin'}, {'rivotril'}, {'clorazep'}, {'tranxen'}, {'diazepam'}, {'valium'}, {'diamox'}, {'diastat'}, {'divalproex'}, {'depakot'}, {'eslicarbazepin'}, {'aptiom'}, {'ethosuximid'}, {'zarontin'}, {'ezogabin'}, {'potiga'}, {'felbam'}, {'felbatol'}, {'gabapentin'}, {'neurontin'}, {'gralis'}, {'lacosamid'}, {'vimpat'}, {{'lamotrigin'}; {'ltg'}}, {'lamict'}, {{'levetiracetam'}; {'lev'}}, {'ige'}, {'keppra'}, {'lorazepam'}, {'ativan'}, {'methsuximid'}, {'celontin'}, {'oxcarbazepin'}, {'trilept'}, {'perampanel'}, {'fycompa'}, {'phenobarbit'}, {'lumin'}, {'phenytoin'}, {'dilantin'}, {'phenytek'}, {'pregabalin'}, {'lyrica'}, {'primidon'}, {'mysolin'}, {'rufinamid'}, {'banzel'}, {'percocet'}, {'stiripentol'}, {'tiagabin'}, {'gabitril'}, {'topiram'}, {'topamax'}, {'qudexi'}, {'trokendi'}, {'valproat'}, {'valproic'}, {' wellbutrin'}, {'depaken'}, {'vigabatrin'}, {'sabril'}, {'zonisamid'}, {'zonegran'}, {'xanax'}. |
| Demographics | Age, sex. |
| International Classification of Diseases (ICD) codes for seizures and epilepsy | 'Number of ICDs', 'convulsions seizures', 'epilepsy and recurrent seizures', 'syncope'. |
| Anti-seizure medications | 'Number of ASMs', 'acetazolamide', 'brivaracetam', 'cannabidiol', 'carbamezapine', 'cenobamate', 'clorazepate', 'eslicarbazepine', 'ethosuximide', 'ezogabine', 'felbamate', 'ketamine', 'lacosamide', 'lamotrigine', 'methsuximide', 'midazolam', 'oxcarbazepine', 'phenobarbital', 'primidone', 'rufinamide', 'tiagabine', 'topiramate', 'valproic acid', 'zonisamide'. |

**Table A.3.** Text features merged based on common medical meaning.

|  |
| --- |
| **{merged bags of words}** |
| {{'histori', 'seizur'}; {'hx', 'seizur'}}, {{'not', 'requir', 'follow', 'up'}; {'not', 'requir', 'followup'}}, {{'follow', 'up', 'as', 'need'}; {'followup', 'as', 'need'}}, {{'seizurefree'}; {'szfree'}; {'seizur', 'free'}; {'sz', 'free'}}, {{'carbamazepin'}; {'cbz'}}, {{'lamotrigin'}; {'ltg'}}, {{'levetiracetam'}; {'lev'}}, {{‘oxcarbazepin’}; {‘oxc’}}, {{'topamax'}; {'tpm'}}, {{'low', 'seizur', 'threshold'}; {'lower', 'seizur', 'threshold'}}, {{'chemo'}; {'chemotherapi'}}, {{'nonepileptiform'}; {'non', 'epileptiform'}}, {{‘no’, ‘epileptiform’, ‘abnorm’}; {‘no’, ‘epileptogen’, ‘abnorm’}}, {{'dejavu'}; {'deja', 'vu'}}, {{'tonic', 'clonic'}; {'tonniclon'}}, {{'recurr', 'sz'}; {'recurr', 'seizur'}}, {{'singl', 'seizur'}; {'singl', 'sz'}}, {'non', 'epilept', 'spell'}; {'nonepilept', 'spell'}}. |

### 2. Results appendix

**Table A.4.** Anti-seizure medications (ASM) distribution for the cohort encounters.

| **ASM, N (%)** | **Encounters**  **(N = 8,415)** |
| --- | --- |
| Acetazolamide | 33 (0.4) |
| Brivaracetam | 234 (2.8) |
| Cannabidiol | 173 (2.1) |
| Carbamezapine | 483 (5.7) |
| Cenobamate | 171 (2.0) |
| Clobazam | 973 (11.6) |
| Clonazepam | 898 (10.7) |
| Clorazepate | 15 (0.2) |
| Diazepam | 453 (5.4) |
| Eslicarbazepine | 225 (2.7) |
| Ethosuximide | 50 (0.6) |
| Ezogabine | 1 (0.1) |
| Felbamate | 46 (0.5) |
| Gabapentin | 541 (6.4) |
| Ketamine | 2 (0.1) |
| Lacosamide | 1,186 (14.1) |
| Lamotrigine | 3,134 (37.2) |
| Levetiracetam | 3,072 (36.5) |
| Lorazepam | 2,131 (25.3) |
| Methsuximide | 5 (0.1) |
| Midazolam | 327 (3.9) |
| Oxcarbazepine | 750 (8.9) |
| Perampanel | 134 (1.6) |
| Phenobarbital | 141 (1.7) |
| Phenytoin | 298 (3.5) |
| Pregabalin | 171 (2.0) |
| Primidone | 62 (0.7) |
| Rufinamide | 72 (0.9) |
| Tiagabine | 25 (0.3) |
| Topiramate | 647 (7.7) |
| Valproic acid | 752 (8.9) |
| Zonisamide | 945 (11.2) |

### A. 2. 1. Hyperparameter tunning

The inverse of the regularization strength in the LR was varied for the values {0.0001, 0.001, 0.01, 0.1, 1.0} and the optimization algorithm (solver) was varied between “liblinear” and “lbfgs”. The following XGBoost parameters were tuned {set of values}: number of trees {100, 150, 200, 250, 300, 350}; maximum tree depth {2, 3, 4, 5}; learning rate {0.01, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1}; gamma, which specifies the minimum loss reduction required to make a split {0, 1, 5}; percentage of features selected to build a tree {30, 40, 50, 60, 70, 80}; and the percentage of notes in train selected to build a tree {80, 90, 100}. For both models, the “warm\_start” parameter was varied between True/False, to reuse the solution of the previous call to fit as initialization (True) or erase the previous solution (False).

**Table A.5.** Hyperparameters selection for the logistic regression (LR) and extreme gradient boosting (XGBoost) models using all features, during training in 5-fold cross validation.

|  |  |  |
| --- | --- | --- |
| **Hyperparameter** | **XGBoost** | **LR** |
| Number of trees | 200 | - |
| Tree depth | 3 |
| Sample of features used to build a tree (%) | 30 |
| Sample of notes used to build a tree (%) | 80 |
| Gamma | 0 |
| Learning rate | 0.09 |
| Warm start | False | True |
| Solver | - | “lbfgs” |
| C | - | 1 |

**Table A.6.** Model performance within the four groups. ASM – anti-seizure medication; ICD – International Classification of Diseases codes for epilepsy and seizures; NPV – negative predictive value; PPV – positive predictive value; TNR – true negative rate; TPR – true positive rate.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ASM+ICD+**   |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | Algorithm | |  |  | | Epilepsy | + | - | Total | TPR | | + | 1793 | 0 | 1793 | 100% | | - | 30 | 0 | 30 | 76.7% | | Total | 1823 | 0 | 1823 | TNR | | PPV | 98.4% | 0% | NPV |  | | **ASM-ICD-**   |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | Algorithm | |  |  | | Epilepsy | + | - | Total | TPR | | + | 0 | 0 | 0 | 100% | | - | 0 | 160 | 160 | 100% | | Total | 0 | 0 | 160 | TNR | | PPV | 100% | 100% | NPV |  | |
| **ASM+ICD-**   |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | Algorithm | |  |  | | Epilepsy | + | - | Total | TPR | | + | 386 | 6 | 392 | 98.5% | | - | 2 | 97 | 99 | 98.0% | | Total | 388 | 103 | 491 | TNR | | PPV | 99.4% | 94.2% | NPV |  | | **ASM-ICD+**   |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | Algorithm | |  |  | | Epilepsy | + | - | Total | TPR | | + | 2 | 0 | 2 | 100% | | - | 0 | 54 | 54 | 100% | | Total | 2 | 54 | 56 | TNR | | PPV | 100% | 100% | NPV |  | |

Performance of the 2-step algorithm within each of the four groups (ASM+ICD+, ASM-ICD-, ASM+ICD, ASM-ICD+). Numbers in green squares indicate true positives (TP, upper left corner) and true negatives (TN, lower right corner); orange squares represent false positives (FP, lower left) and false negatives (FN, upper right). Positive predictive values PPV = TP/(TP+FP) and the true positive rate TPR = TP/(TP+FN) (aka sensitivity) are shown in yellow; negative predictive value NPV = TN/(TN+FN) and the true negative rate (aka specificity) are shown in blue. Note that step 1 of the model applies to the “easy” cases (ASM+ICD+: classify as +, ASM-ICD-: classify as -); step 2 of the model is the logistic regression model (see main text). The overall error rates within each group are: ASM+ICD+: 30/1823 = 1.64%; ASM-ICD-: 0/160 = 0%; ASM+ICD-: 8/491 = 1.63%; ASM-ICD+: 0/56 = 0%.

**Table A.7.** Average performance [95% confidence intervals] for logistic regression (LR) and extreme gradient boosting (XGBoost) models using all features in the full test set.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Classes** | **Model** | **AUROC** | **AUPRC** | **F1-score** | **Recall** | **Precision** | **Specificity** |
| Macro average | LR | 1.00  [0.99-1.00] | 1.00  [1.00-1.00] | 0.96  [0.95-0.98] | 0.97  [0.96-0.98] | 0.96  [0.94-0.97] | 0.97  [0.96-0.98] |
| XGBoost | 0.96  [0.95-0.97] | 0.96  [0.95-0.98] | 0.95  [0.94-0.97] | 0.96  [0.95-0.98] |
| Epilepsy | LR | 1.00  [0.99-1.00] | 1.00  [0.99-1.00] | 0.99  [0.99-0.99] | 0.99  [0.98-0.99] | 0.99  [0.99-1.00] | 0.96  [0.94-0.98] |
| XGBoost | 0.99  [0.98-0.99] | 0.98  [0.98-0.99] | 0.95  [0.93-0.97] |
| No epilepsy | LR | 1.00  [0.99-1.00] | 0.97  [0.95-0.99] | 0.94  [0.92-0.96] | 0.96  [0.94-0.98] | 0.93  [0.90-0.95] | 0.99  [0.98-0.99] |
| XGBoost | 0.93  [0.91-0.95] | 0.95  [0.93-0.97] | 0.90  [0.87-0.93] | 0.98  [0.98-0.99] |

AUROC – area under the receiver operating characteristic; AUPRC – area under the precision-recall curve; LR – logistic regression, XGBoost – extreme gradient boosting model.

**Figure A.1.** Areas under the (a) receiver operating characteristic and (b) precision-recall curves, for the logistic regression model using all features in the full test set. ‘YES’/ ‘NO’ refer to positive/negative diagnosis of epilepsy.

Diagram

Description automatically generated with medium confidence A picture containing graphical user interface

Description automatically generated

1. (b)

**Figure A.2.** Areas under the (a) receiver operating characteristic and (b) precision-recall curves, for the logistic regression model using all features in the test set including patients with International Classification of Diseases codes for seizures or anti-seizure medications. ‘YES’/ ‘NO’ refer to positive/negative diagnosis of epilepsy.

Diagram

Description automatically generated with medium confidence A picture containing table

Description automatically generated

1. (b)

**Figure A.3.** Confusion matrices for the logistic regression model (a) (d) normalized by recall, (b) (e) normalized by precision and (c) (f) without normalization, for the full test set (a)-(c) and for the test set including patients with International Classification of Diseases codes for seizures or anti-seizure medications (d)-(f). ‘YES’/ ‘NO’ refer to positive/negative diagnosis of epilepsy.

Chart, bar chart

Description automatically generated Chart, bar chart

Description automatically generated Chart, waterfall chart

Description automatically generated

1. (b) (c)

Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generatedChart

Description automatically generated

1. (e) (f)

**Figure A.4.** Top 20 most important features of the extreme gradient boosting model using the set with all features. SHAP – SHapley Additive exPlanations. For each bag of words in brackets, words are in their stemmed form, and they all appear in at least one sentence of the patient visit note. Positive SHAP value corresponds to positive diagnosis of epilepsy. ASMs – anti-seizure medications; ICDs – International Classification of Diseases codes for seizures and epilepsy.

Chart

Description automatically generated with low confidence