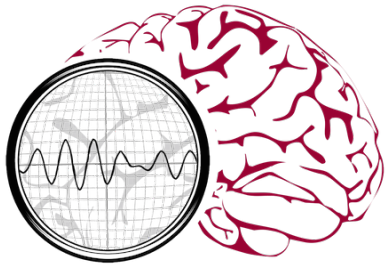


The 2020 IEEE Signal Processing in Medicine and Biology Symposium

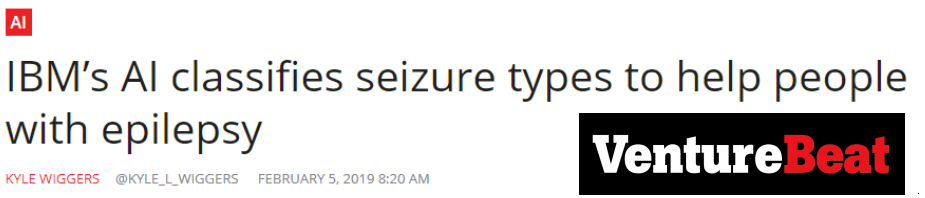
Oral Presentation

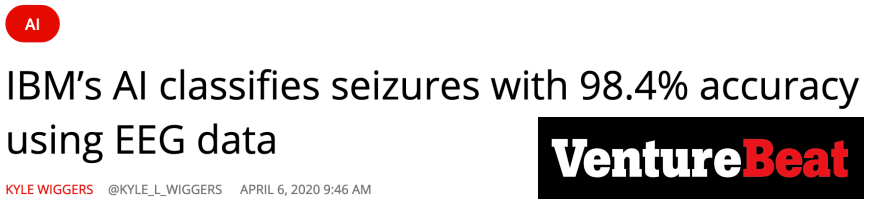
IEEE

Seizure Type Classification Using EEG signals and Machine Learning: Setting a Benchmark

Subhrajit Roy, Umar Asif, Jianbin Tang, Stefan Harrer

Contact: jbtang@au1.ibm.com





IBM Research

Dataset: TUH EEG Seizure Corpus

- Most comprehensive publicly available EEG dataset: >30000 sessions since 2002, ~2500 new sessions per year
- Each session contains: annotated EEG signal data and de-identified medical reports written by certified neurologists.

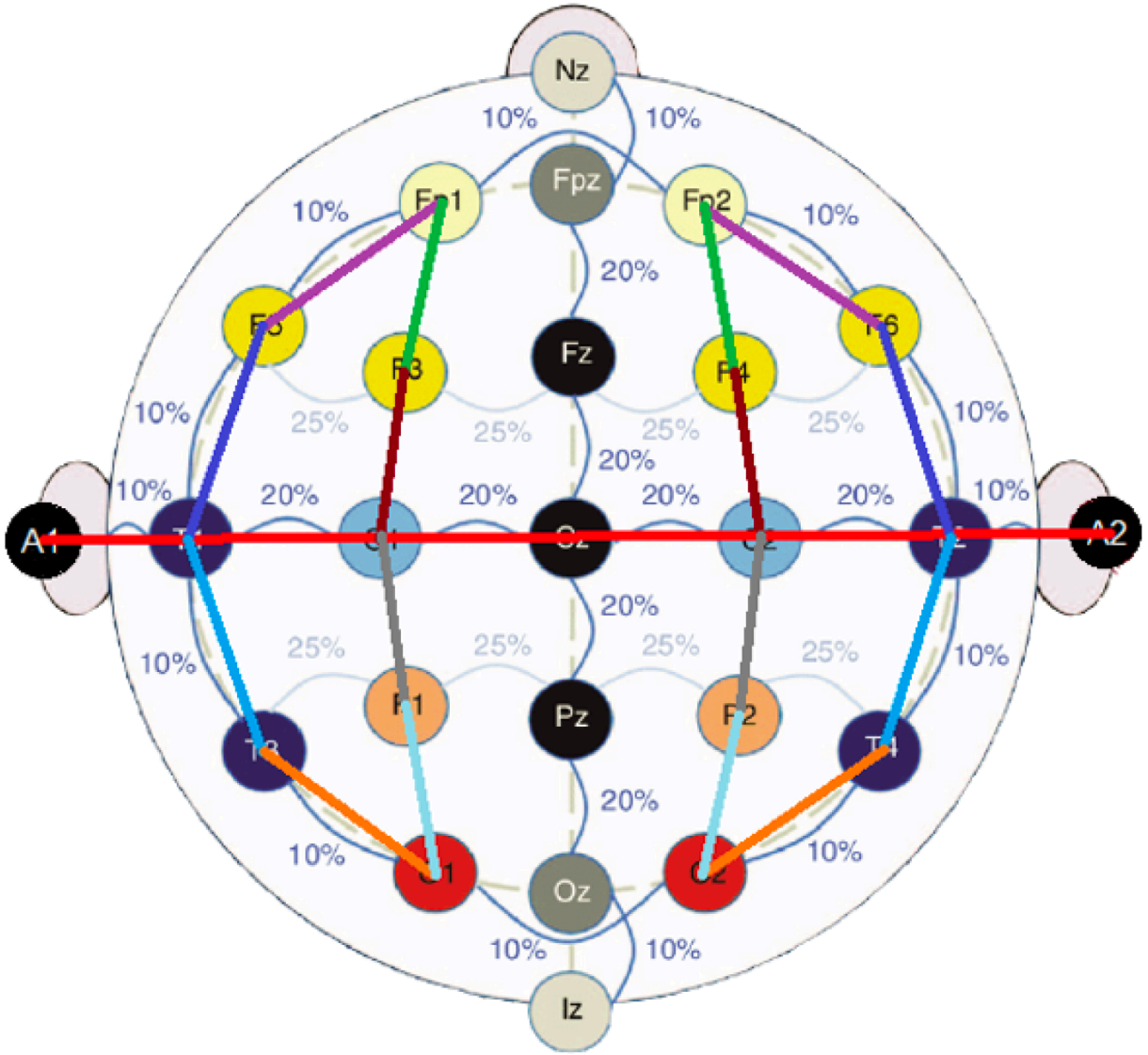
Table 1. Seizure Type Statistics for v1.4.0

Seizure Type	Seizure Number	Duration (Seconds)	Patient Number
Focal Non-Specific (FNSZ)	992	73466	109
Generalized Non-Specific (GNSZ)	415	34348	44
Complex Partial (CPSZ)	342	33088	34
Absence (ABSZ)	99	852	13
Tonic (TNSZ)	67	1271	2
Tonic Clonic (TCSZ)	50	5630	11
Simple Partial (SPSZ)	44	1534	2
Myoclonic (MYSZ)	3	1312	2

Table 2. Seizure Type Statistics for v1.5.2

Seizure Type	Seizure Number	Duration (Seconds)	Patient Number
Focal Non-Specific (FNSZ)	1836	121139	150
Generalized Non-Specific (GNSZ)	583	59717	81
Complex Partial (CPSZ)	367	36321	41
Absence (ABSZ)	99	852	12
Tonic (TNSZ)	62	1204	3
Tonic Clonic (TCSZ)	48	5548	14
Simple Partial (SPSZ)	52	2146	3
Myoclonic (MYSZ)	3	1312	2

Temporal Central Parasagittal (TCP) montage*



FP1—F7 FP2—F8

F7—T3 F8—T4

T3—T5 T4—T6

T5—O1 T6—O2

A1—T3 T4—A2

T3—C3 C4—T4

C3—CZ CZ—C4

FP1—F3 FP2—F4

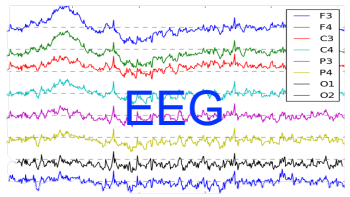
F3—C3 F4—C4

C3—P3 C4—P4

P3—O1 P4—O2

*Vanabelle P, De Handschutter P, El Tahry R, et al. Epileptic seizure detection using EEG signals and extreme gradient boosting[J]. The Journal of Biomedical Research, 2020, 34(3): 226-237.

Data Pre-Processing



Resample to 250Hz

TCP Montage

Method 1: Generate FFT Images[1]

FFT(1~ f_{max} Hz) → Log10(abs(F))

Method 2: Generate FFT Correlation Coef[2]

FFT(1~ f_{max} Hz) → Correlation Coefficient Matrix+ Eigenvalues

50 combinations:
Window = {1, 2, 4, 8, 16} Seconds
Overlap = {0.5W, 0.75W} Seconds
f_{max} = {12, 24, 48, 64, 96} Hz

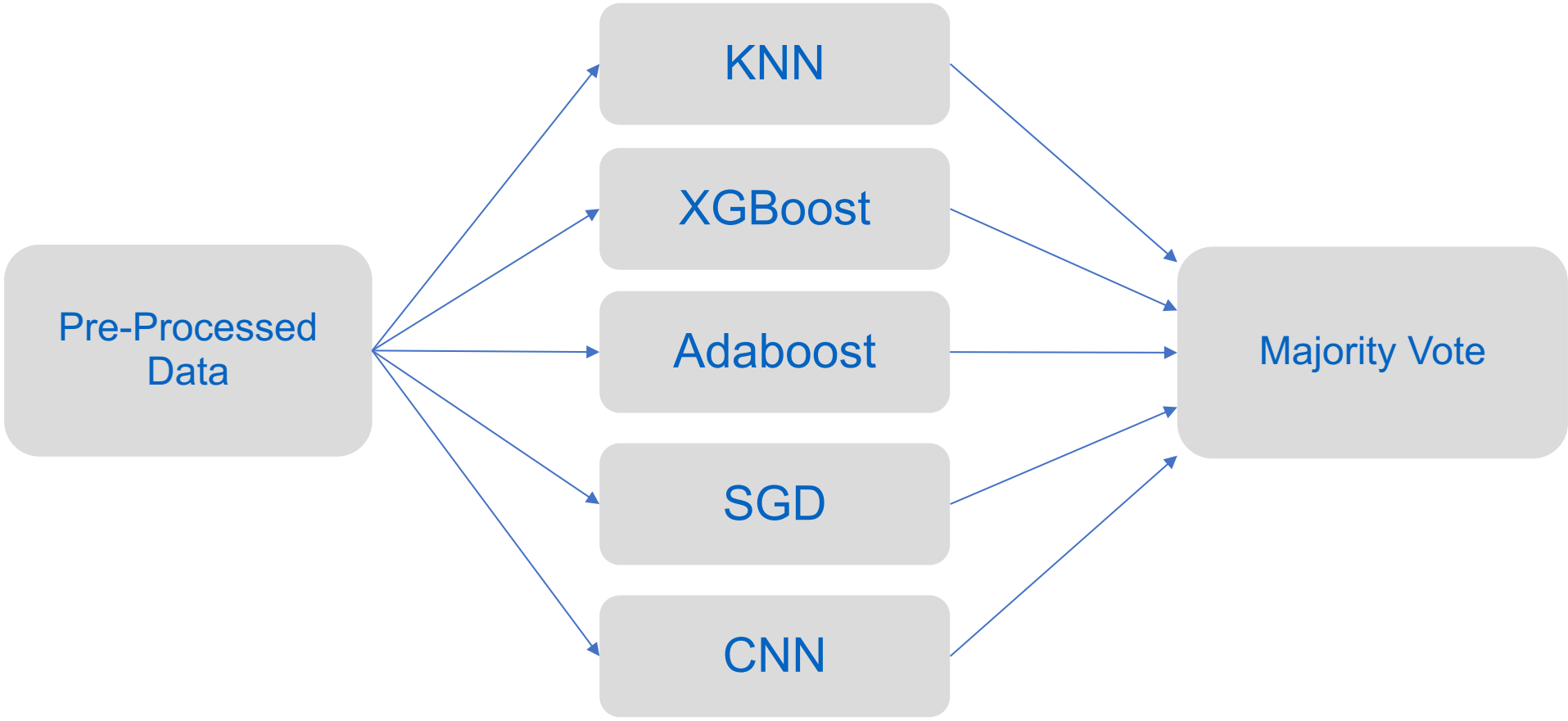
Open source: https://github.com/IBM/seizure-type-classification-tuh

master · 1 branch · 0 tags · Go to file · Code

Jianbin-IBM add utils	c8e7a22 on Aug 25	7 commits
data_preparation	add utils	3 months ago
preprocess	.gitignore fix	4 months ago
utils	add utils	3 months ago
.gitignore	add gitignore	4 months ago
LICENSE	Initial commit	4 months ago
README.md	add utils	3 months ago

[1] Y. Paul, "Various epileptic seizure detection techniques using biomedical signals: a review," Brain informatics, vol. 5, no. 2, p. 6, 2018.
[2] K. Schindler, H. Leung, C. E. Elger, and K. Lehnertz, "Assessing seizure dynamics by analyzing the correlation structure of multichannel intracranial eeg," Brain, vol. 130, no. 1, pp. 65–77, 2007.

Baseline Machine Learning Framework



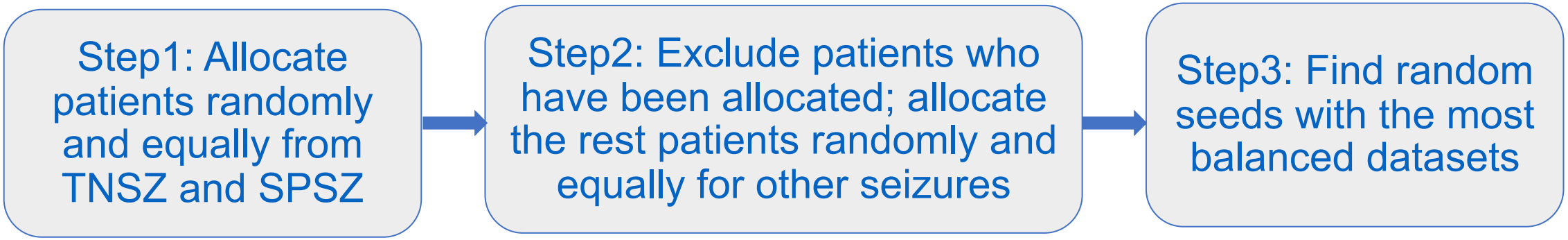
Weight F1 has been used to evaluate the multi-classification performance

$$Weighted\_F1 = \sum_{n=1}^{7} \frac{\alpha_n \times F1_n}{7}$$

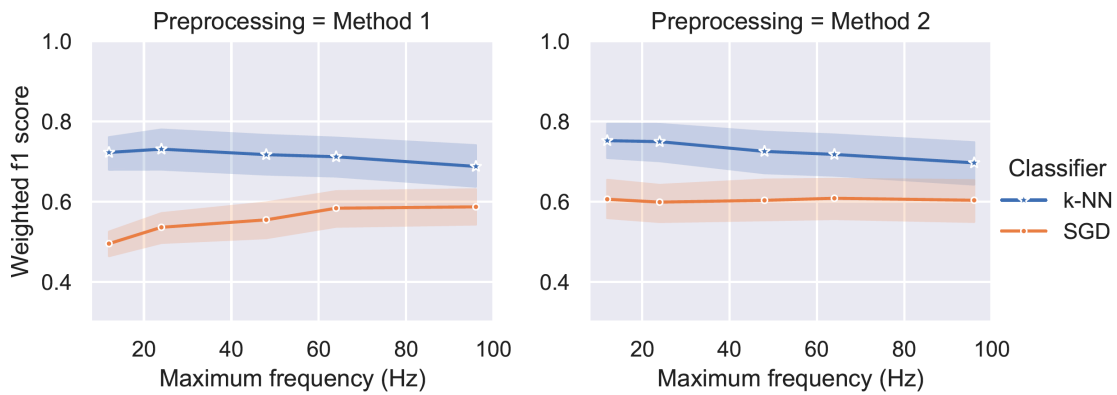
$$\alpha_n = \frac{Number\ of\ Seizure\ Type\ n}{Total\ Seizure\ Number}$$

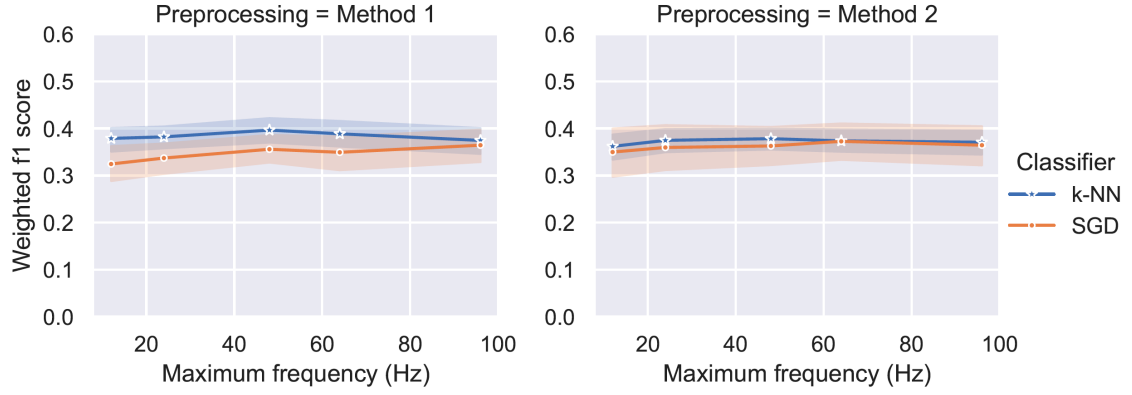
Cross validation Scheme

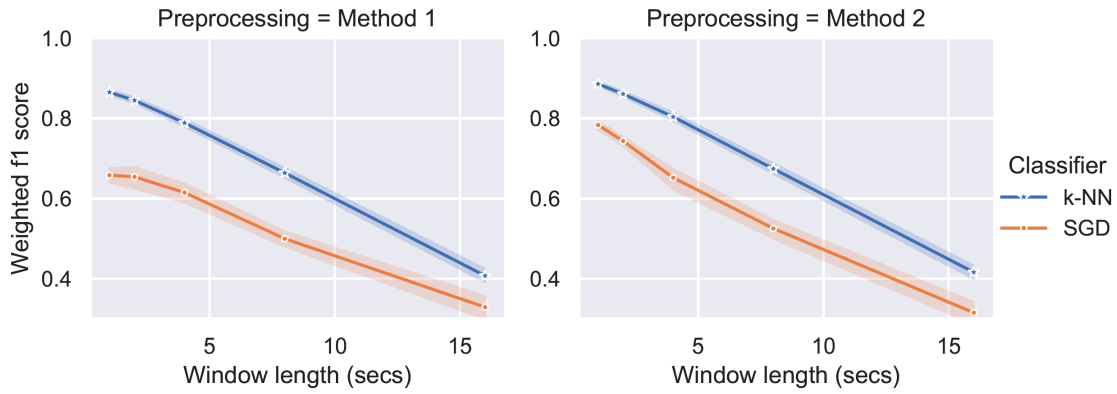
- 5 folds seizure wise cross validation for v1.4.0
  - Randomly divide seizures from each type into 5 groups
  - 1 group for testing, the rest 4 groups for training
  - Same patients' data will be used for training and testing
- 3 folds patient-wise cross validation for v1.5.2
  - Balanced Patient Number and Seizure Number for each group

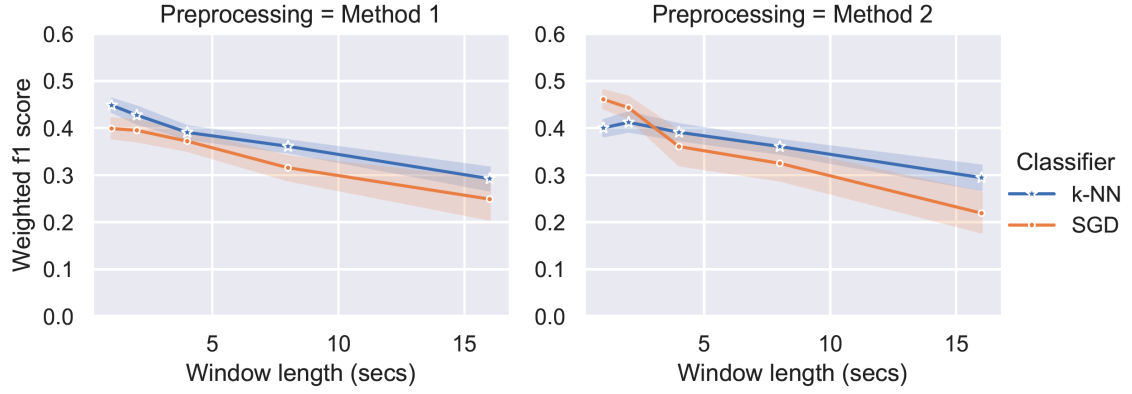


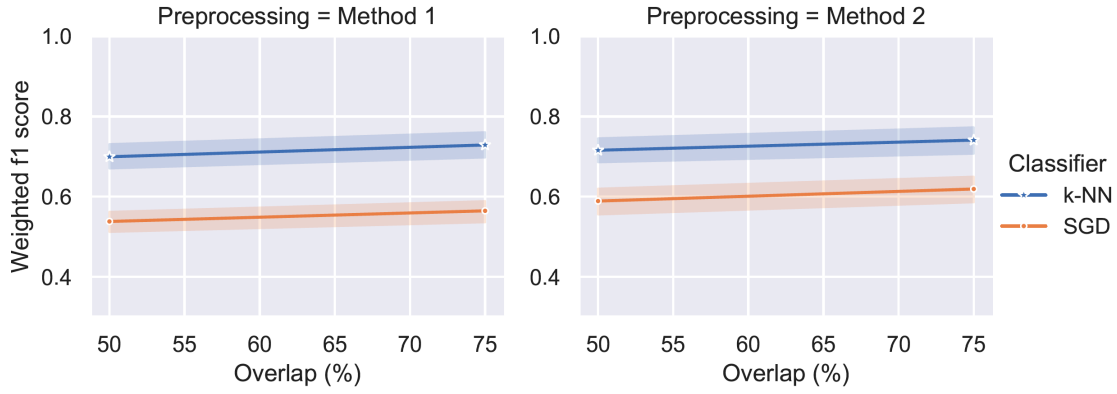
Seizure Num	Group 1	Group 2	Group 3
Train	2031	2033	2030
Val	1016	1014	1017

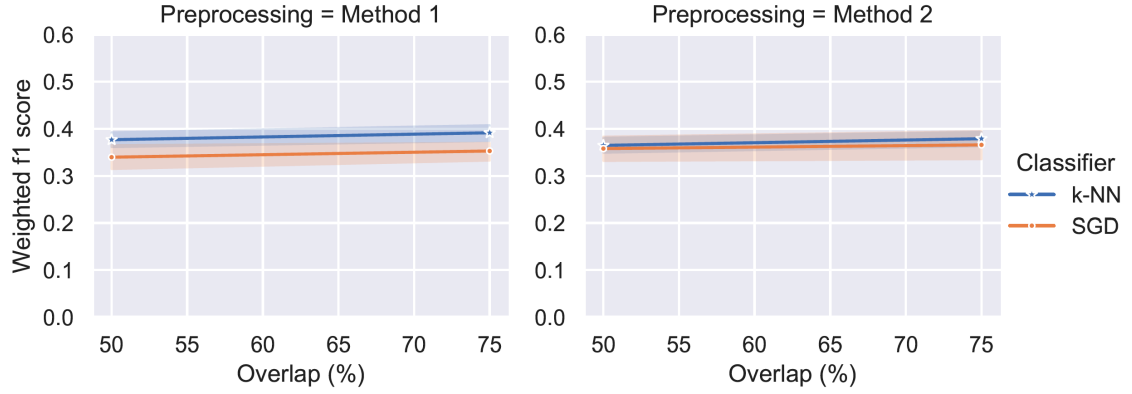












v1.4.0

v1.5.2

How the weighted-F1 score varies with f_{max} (top row), Window Length (middle row), and Overlap (bottom row) for both pre-processing techniques on k-NN and SGD classifier

v1.4.0

5 folds seizure wise cross validation

	f_{max}	W_l	O	$k-NN$	SGD	$XGBoost$	CNN
Method 1	48	1	$0.75W_l$	**0.884**	0.695	0.817	0.714
	24	1	$0.75W_l$	0.883	0.621	0.844	0.722
	96	1	$0.75W_l$	0.880	0.724	0.745	0.718
	24	1	$0.5W_l$	0.879	0.604	0.766	0.713
Method 2	48	1	$0.75W_l$	**0.901**	0.807	0.851	*NA*
	24	1	$0.75W_l$	0.900	0.783	0.858	*NA*
	24	1	$0.5W_l$	0.895	0.752	0.819	*NA*
	96	1	$0.75W_l$	0.890	0.806	0.866	*NA*

v1.5.2

3 folds patient wise cross validation

	f_{max}	W_l	O	$k-NN$	SGD	$XGBoost$	CNN
Method 1	96	1	$0.75W_l$	0.466	0.432	**0.561**	0.524
	24	1	$0.75W_l$	0.437	0.384	0.559	0.530
	48	1	$0.75W_l$	0.467	0.407	0.526	0.525
	24	1	$0.5W_l$	0.423	0.390	0.512	0.504
Method 2	48	1	$0.75W_l$	0.401	0.469	**0.542**	*NA*
	96	1	$0.75W_l$	0.418	0.459	0.535	*NA*
	24	1	$0.5W_l$	0.392	0.452	0.530	*NA*
	24	1	$0.75W_l$	0.412	0.462	0.524	*NA*

Thank you for listening!

Jianbin Tang (jbtang@au1.ibm.com)