A Comparative Study Between Classical Feature Engineering and RNNs for Seizure Detection in Imbalanced Data

Mashael Alzaid

Princess Noura Bint Abdulrahman University

Riyadh, Saudi Arabia

The 2020 IEEE Signal Processing in Medicine and Biology Symposium

"Epilepsy is a neurological disorder that causes a sudden and repeated rush of electrical activity in the brain known as seizures."

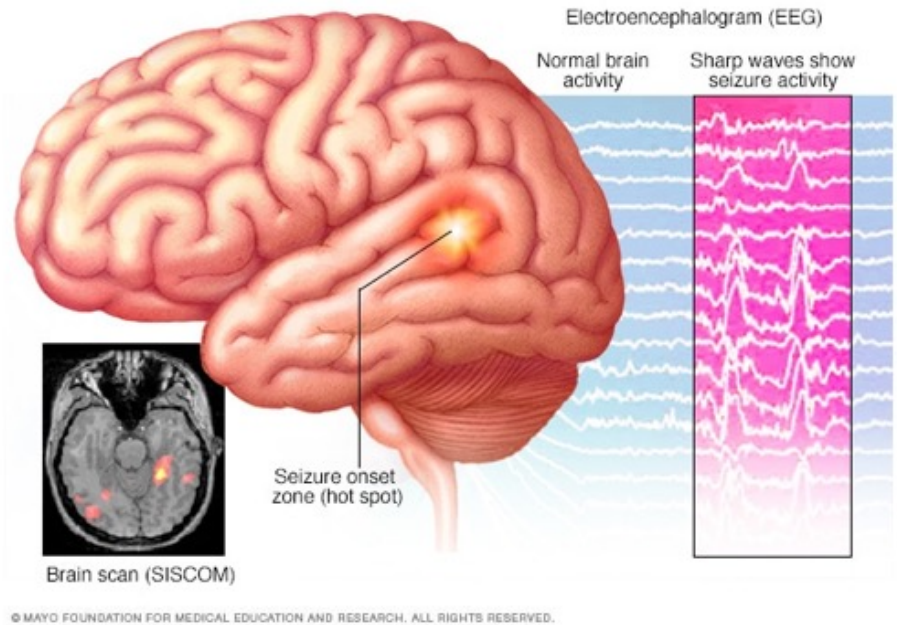


Image: https://www.mayoclinic.org/ar/diseases-conditions/epilepsy/symptoms-causes/syc-20350093

Why is it important?

50 M

Are diagnosed with epilepsy

1:50

Person will be diagnosed with epilepsy at some point in their life .

10%

of people worldwide will have one seizure during their lifetime

Enhancing Seizure detection using ML

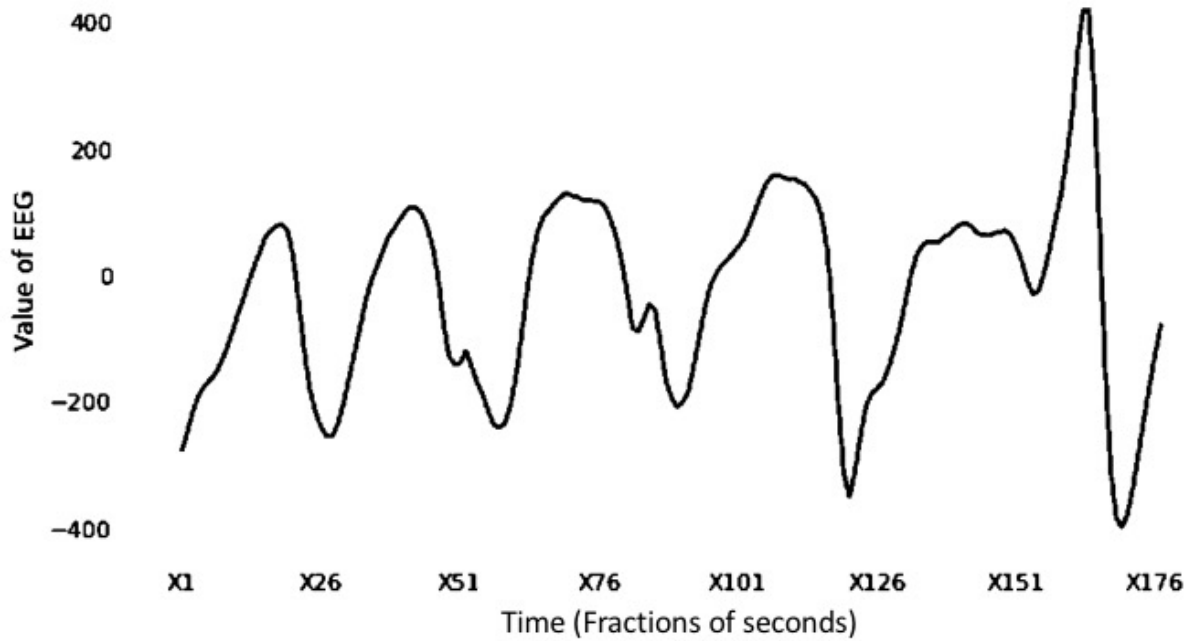
Was the event a seizure or something that looks like a seizure?

(Migraine headaches, muscle cramps, and sleep disturbances ... etc**)**

Which models, and settings is best to detect seizures in imbalanced datasets?

The visual analysis of EEG recordings is a time-consuming and error-prone procedure

Patient 874's EEG recoding during a seizure event



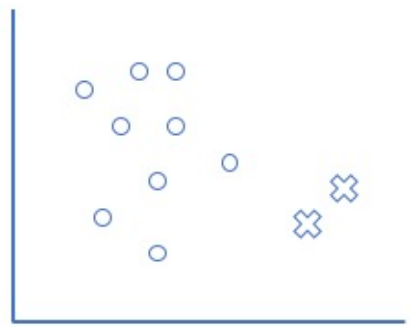
EXISTING WORK

ML in seizure detection

In some of the work the authors used deep learning approach and scored 100% none the less the performance would vary sharply from one dataset to the other, that doesn't guarantee that the model can be generalized and can perform well in deployment stage. Furthermore, the sensitivity in this approach scored 100% however the amount of improvement comparing to other approaches included in the experiment is low. Others used classical ML algorithm such as SVM, DT, and KNN and a manual feature extraction to predict seizure

ML in imbalanced datasets

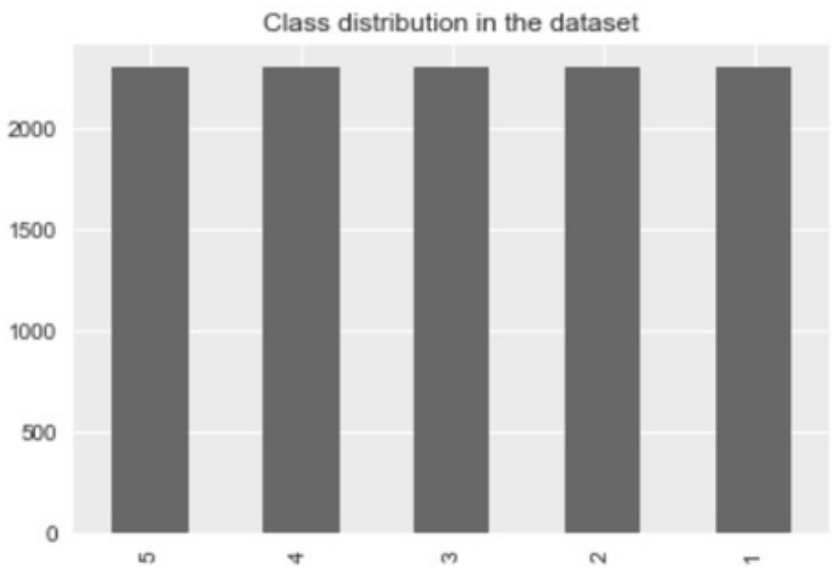
Researchers showed SVM is sensitive to imbalanced datasets and worked on improving its accuracy using hybrid techniques however the sensitivity in these algorithms fluctuated sharply from one dataset to the other and traditional sampling techniques would outperform them in other. This could mean that the best solution for handling imbalanced dataset differs from one domain to the other and from one dataset to the other



Dealing with imbalanced medical datasets

Researchers have developed sampling and algorithmic solutions to deal with imbalanced datasets however the algorithms either would not be able to outperform other approaches such in certain datasets, other researchers developed a combination of sampling with an algorithm such as Naive Bayes to handle the imbalance problem nonetheless the amount of improvement is very little, other experimented on traditional approaches such as SMOTE, under and over sampling or combination of these methods however they don't consider the loss in data or the drop in data quality

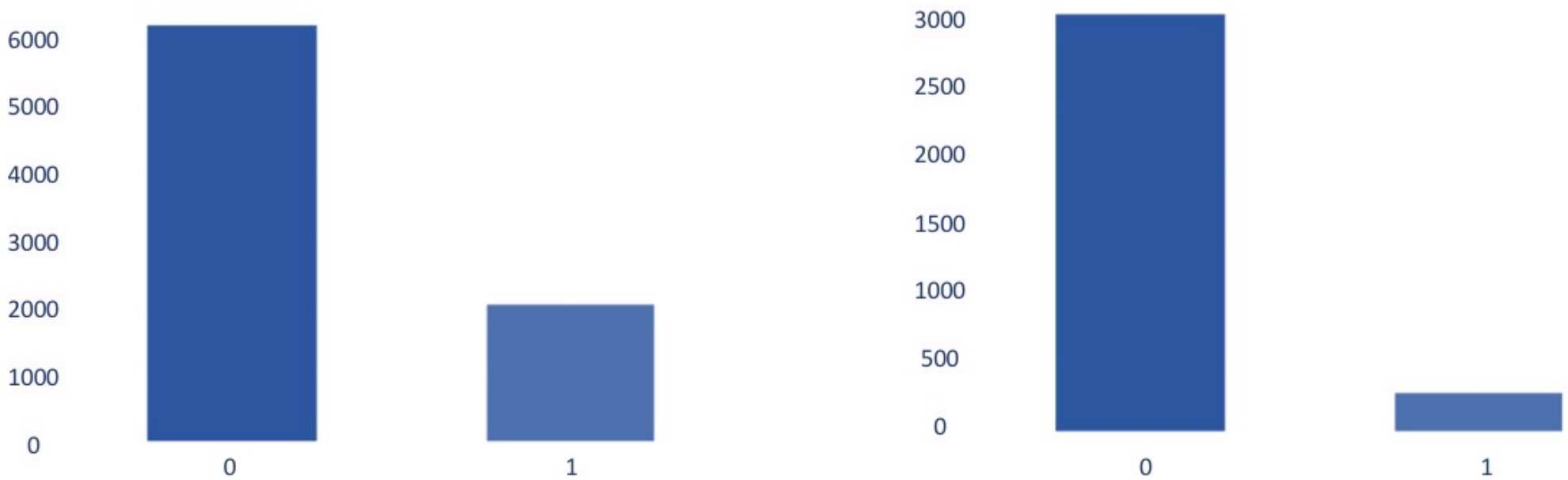
Targeting the population of interest



Original population

Target population

Data Distribution in Training and Testing Sets



Training (left) and testing (right) sets before ratio balancing.

Quality and Safety of ML System in the medical field

ML System perform better on balanced datasets

•Haixiang G, Yijing L, Shang J, *et al*. Learning from class-imbalanced data: review of methods and applications. Expert Syst Appl 2017;**73**:220-39.doi:10.1016/j.eswa.2016.12.035

Unbalanced data may be rebalanced, and without correction the resulting system tends to over-estimate the rare event.

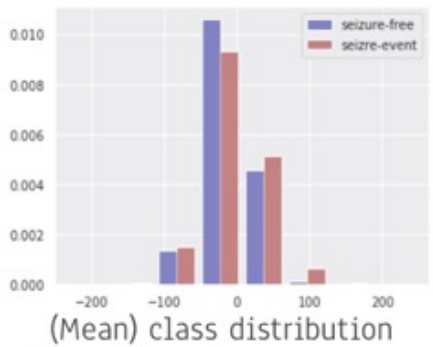
•Lawrence CSS,Storkey AJ. When Training and Test Sets are Different: Characterising Learning Transfer. In: Lawrence CSS, ed. *Dataset shift in machine learning*. MIT Press, 2013: 3–28.

Dummy & Baseline model

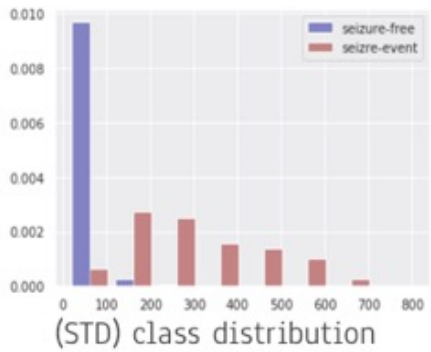
Approach	Accuracy	Precision	F1	Recall	AUC
Dummy Classifier	0.69	0.12	0.15	0.24	0.49
BRR on original data	0.8	0.9	0.04	0.02	0.5

Performance of dummy classifier vs. baseline model

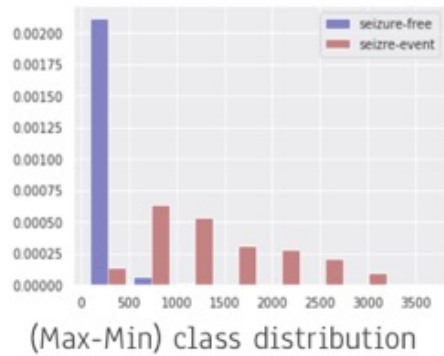
Extracting Statistical Features



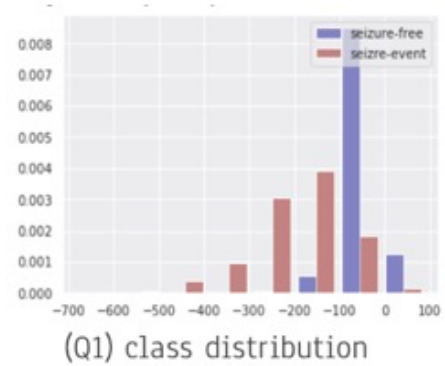
(Mean) class distribution



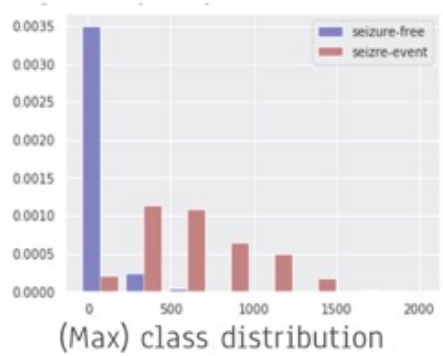
(STD) class distribution



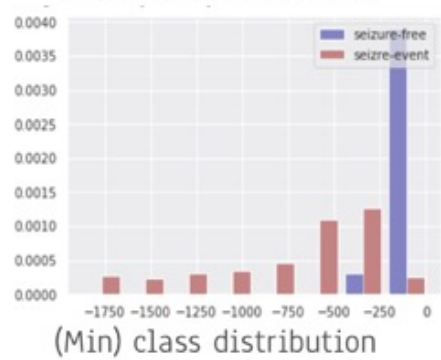
(Max-Min) class distribution



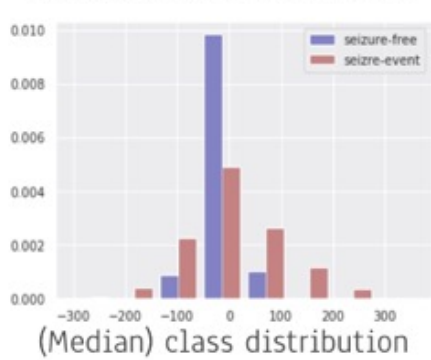
(Q1) class distribution



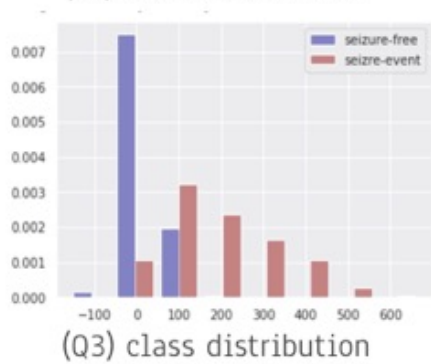
(Max) class distribution



(Min) class distribution



(Median) class distribution



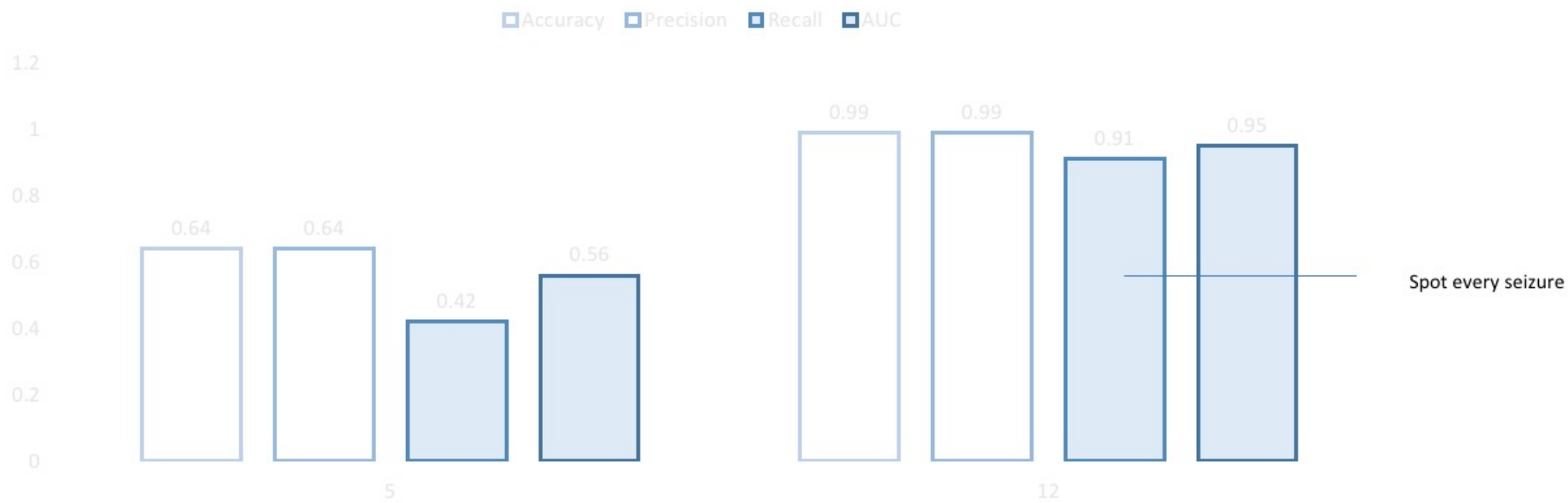
(Q3) class distribution

Extracted Statistical Features

Target	Statistics	p-value
Mean	4.761	0.029
Standard Deviation	5308.046	0
Max	5025.22	0
Min	5152.842	0
Max-Min	5269.803	0
Q1	4374.825	0
Median	17.76	0
Q3	4004.397	0

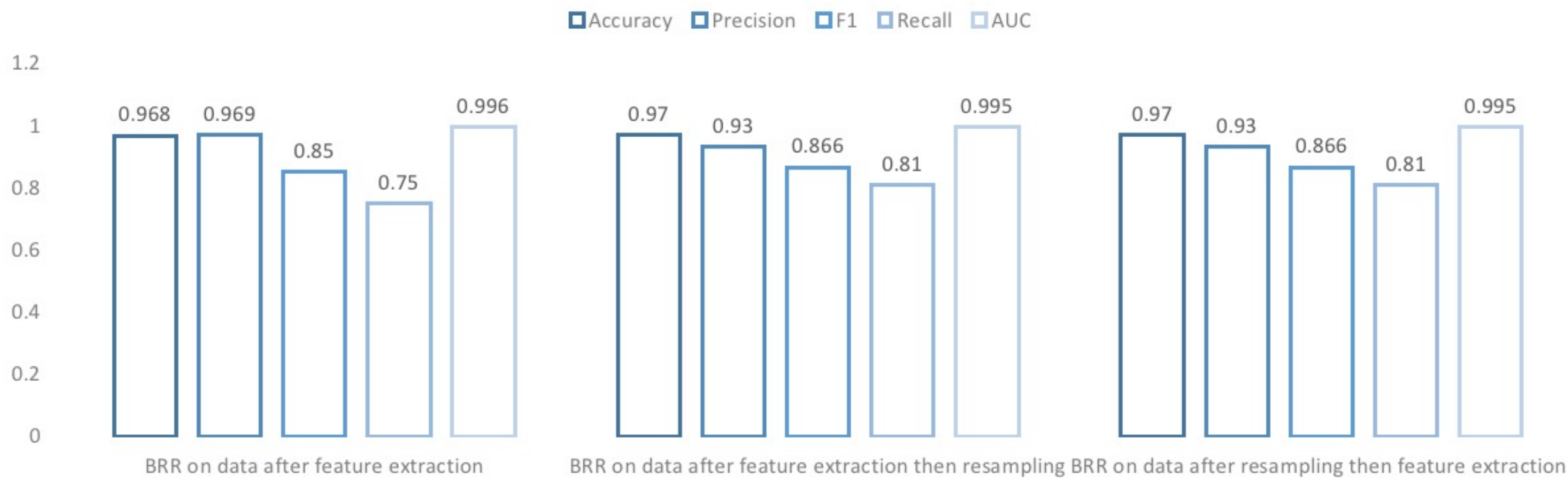
Kruskal Walli test

Resolving possible source of data leakage



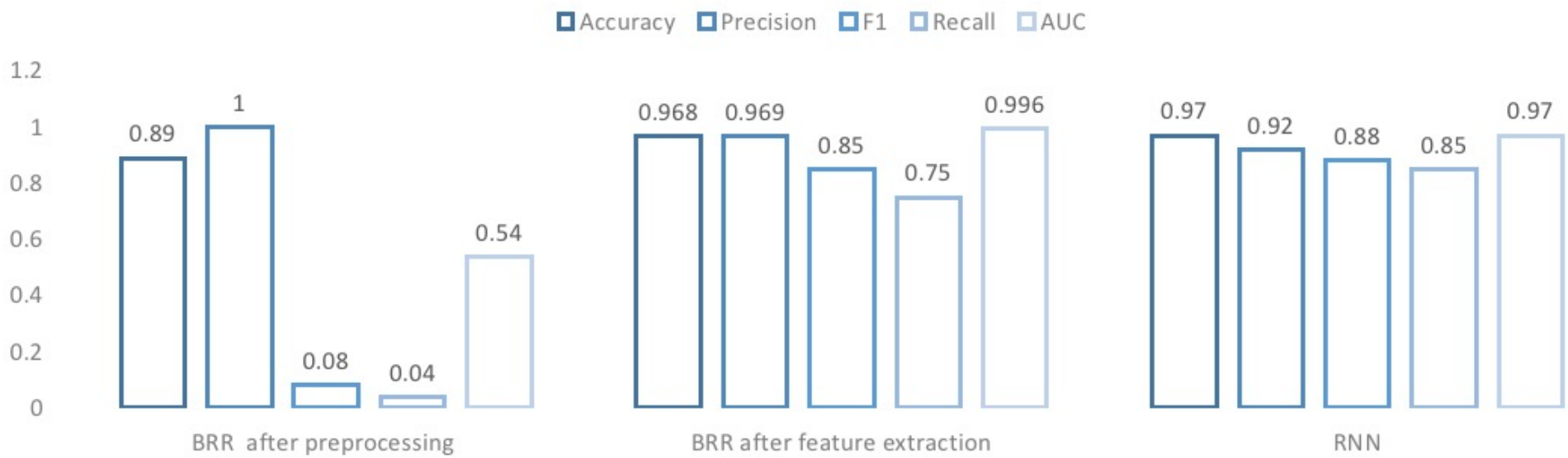
Performance after addressing data leakage

Resampling & Feature extraction Ordering

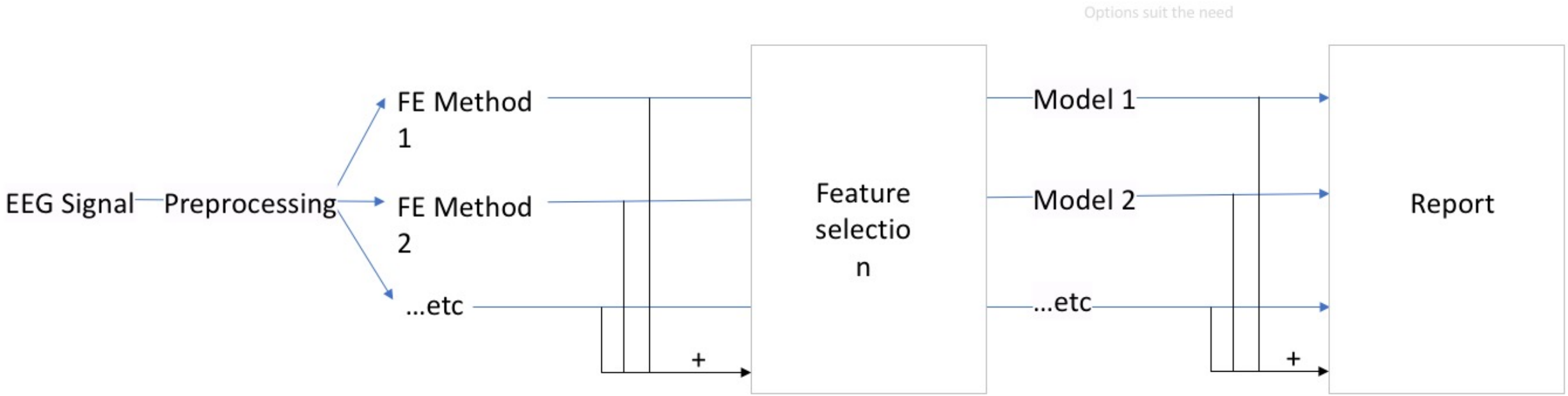


Performance of models with different orderings.

Classical feature engineering Vs RNN



Performance of classical ML models and RNN.


Options suit the need
EEG Signal
Preprocessing
FE Method 1
FE Method 2
…etc
+
Feature selectio n
Model 1
Model 2
…etc
+
Report

Dr. Khalid Abdul
neurologist

AI4Epilepsy

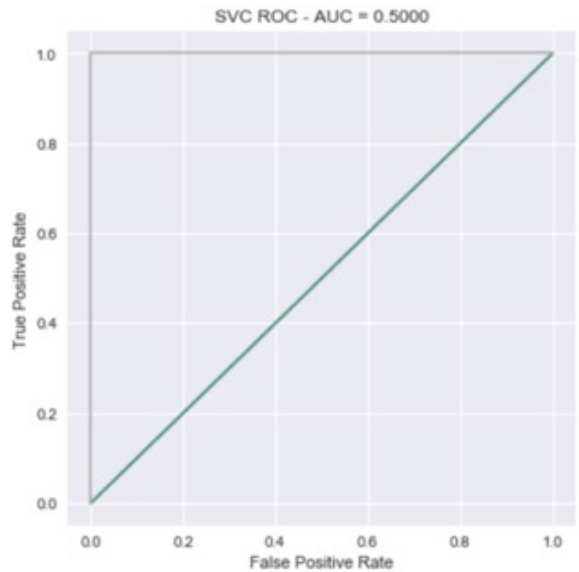
A tool that utalizes AI in detecting epileptic seizures in EEG recordings

Start

Note

```
Accuracy Score : 0.7976811594202898
Precision Score : 0.7976811594202898
F1 Score : 0.7976811594202898
Recall Score : 0.0
AUC of ROC Curve: 0.5
```

	precision	recall	f1-score	support
0	0.80	1.00	0.89	2752
1	0.00	0.00	0.00	698
accuracy			0.80	3450
macro avg	0.40	0.50	0.44	3450
weighted avg	0.64	0.80	0.71	3450

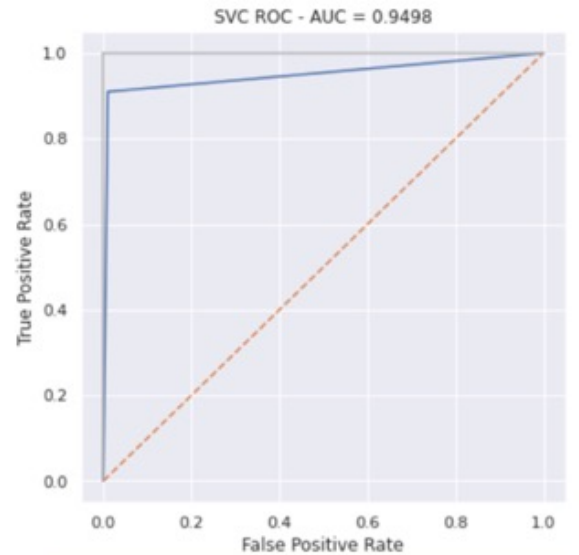



```
print(sklearn.__version__)
```

```
0.21.3
```

```
Accuracy Score : 0.9736231884057971
Precision Score : 0.9736231884057971
F1 Score : 0.9736231884057971
Recall Score : 0.9097421203438395
AUC of ROC Curve: 0.949783850869594
```

	precision	recall	f1-score	support
0	0.98	0.99	0.98	2752
1	0.96	0.91	0.93	698
accuracy			0.97	3450
macro avg	0.97	0.95	0.96	3450
weighted avg	0.97	0.97	0.97	3450

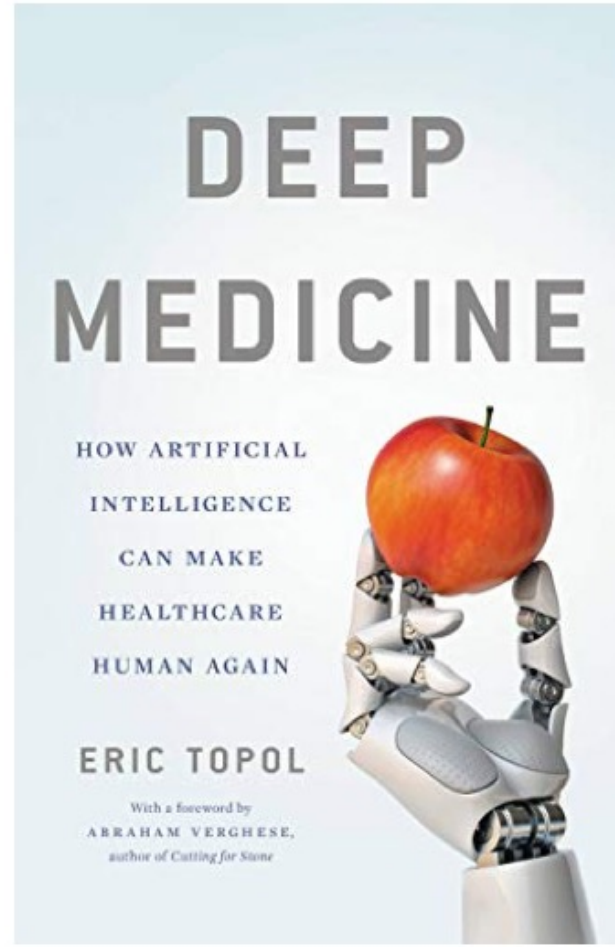



```
print(sklearn.__version__)
```

```
0.22.2.post1
```

A Day in the Life of Someone with Epilepsy

https://brainsandbodiesblog.com/2017/10/08/a-day-in-the-life-of-someone-with-epilepsy/



https://www.amazon.com/Deep-Medicine-Artificial-Intelligence-Healthcare-ebook/dp/B07FMHFGLT

Special Thank you

Dr. Nazek Alturkey
Professor @PNU
Saudi Arabia

Dr. Graham Healy
Professor @DCU
Ireland

Anka Chan
Neuroscience Student At
University of Alberta
Co-founder @Hempact Inc.
Canada

Eng. Haroun Moula
CEO @Epilert
Tunisia

جامعة الأميرة نورة بنت عبدالرحمن
Princess Nourah Bint Abdulrahman University

AEC
شركة الإلكترونيات المتقدمة
Advanced Electronics Company

DCU

Let's stay in touch

Mashael Alzaid

Mashaelalzaid@gmail.com