

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

M. Alzaid

Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia
Masha'alzaid@gmail.com

Abstract— Epilepsy is one of the most common neurological diseases worldwide, defined as a number of seizures in the brain that affect the person's quality of life and her/his ability to perform regular activities. Epilepsy is diagnosed in several ways, one of which is a test known as EEG. The process of examining brain activity is a long and error-prone process. Researchers developed machine learning algorithms to identify epileptic seizures through available datasets. This study aims to compare the performance of a Machine Learning ML model with classical feature engineering approach and RNN for seizure detection in imbalanced dataset. In the experiment, feature engineering improved the recall dramatically in the classical pipeline and slightly improved it in RNN where there is no (manual) feature engineering. The AUC also had increased after feature engineering in classical pipeline and RNN, exceeding 95%.

it was clear that the feature engineering improved the performance dramatically in the classical pipeline, and that the order of the feature engineering and resampling (downsampling) didn't affect the performance, and although applying a classical ML model on manually extracted features and RNN gave approximately close results, the recall which spots every seizure was better in RNN comparing to the classical approach.

I. INTRODUCTION

Epilepsy is a neurological disorder that is defined as having unprovoked, recurrent and sudden rush of electrical activity in the brain known as seizures. Epilepsy affects approximately 50 million people around the world, 70% of whom could live seizure-free if diagnosed and treated properly[1]. Nonetheless, diagnosing epilepsy is difficult for many reasons: first, other conditions, such as fainting, migraines and panic attacks, can cause similar symptoms[2]. Second, in many societies worldwide, being diagnosed with epilepsy is surrounded by stigma, fear and discrimination. Third, the fact that diagnosing a person with epilepsy is a long and error-prone process. A practitioner may check the patient's brain activity through a test called electroencephalogram (EEG) which records the brain signals and is used to diagnose and monitor conditions such as epilepsy. [3]

Machine learning was used to detect epileptic seizure with the aim to reduce error rate and boost practitioners' performance.

Training Machine Learning models can suggest misleading decisions and degrading the performance of a classifier when applied on imbalanced data where the class distribution is significantly skewed and represents certain groups more than the others. Over the years, researchers have developed techniques to handle imbalanced data sets, such as sampling, algorithm techniques and hybrid. In some of the work, the authors argued that certain class imbalance learning methods could lead to information loss and overlapping data and hence may not be the best choice to handle the problem. Others investigated modifying algorithms which were either successful when applied on imbalanced datasets or sensitive to it and therefore produce suboptimal models.

This paper focuses on comparing the performance of a Machine Learning ML model with classical feature engineering approach and RNN for seizure detection in imbalanced dataset.

II. DATA SOURCE AND PREPROCESSING

The data was collected from a repository for epileptic seizure EEG data [17]. The dataset consists of 23 seconds of EEG recordings for every patient of the 500 patients, which makes a total of 11500 records and 179 features. The class distribution of the original dataset is balanced. Nonetheless, converting relabeling the classes

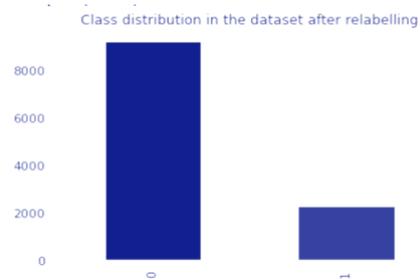


Figure 1. Class distribution after relabeling classes.

(1-seizure and 0-seizure-free), caused a significant imbalanced distribution between the new classes.

Data Leakage

Leakage in data mining is “the introduction of information about the target of a data mining problem, which should not be legitimately available to mine from”[18]. Leakage can make the model's performance look promising yet can result in overfitting and failing to generalize. The possible source of data leakage in this work is known as Leaky Validation Strategy according to [19] which is derived from not distinguishing training data from validation data, as described in the data collection section. Knowing that there are 500 patients in the dataset, and that every patient has 23 rows (or seconds) in the dataset, building a model without identifying which data belongs to which patient could result in data leakage, and therefore, the aim of this part is to identify which 23 records belong to which patient to be able to divide the patients into 1 of two sets (Test or Train and not both).

Class ratio in training and testing sets

After relabeling, the class distribution ratio was (1:3) in training set and (1:10) in testing set, in this part of the experiment, the ration was balanced between seizure and non-seizure events in training and testing sets. Therefore approximately 43 patients with seizure were moved from training to the testing set. With that, Training set consists of 64% of the total data with 7324 records, and the testing set has 36% of the total set with 4176 records.

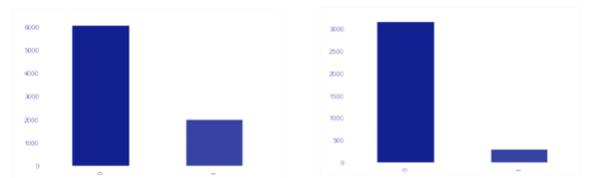


Figure 2. Training (left) and testing (right) sets before ratio balancing.

III. Feature Engineering

This work covers Statistical and Fast Fourier Transform features. FFT is an algorithm that transforms signals into its frequency representation [21]. The FFT dataset was merged with 8 statistical features which were derived from the 178 features in the raw data for each record: Mean, Standard Deviation, Minimum value, maximum value, maximum value - minimum value, First Quartile, 2nd (Median) and third quartile.

A Kruskal Wallis Test[20], which tests whether the population median of all of the groups are equal (The

SciPy community, 2019), was applied on the dataset with 5% level of significance, comparing the distribution of each of the extracted features then applied on seizure versus seizure-free populations for each feature. The null hypothesis is rejected for all the features, meaning different distributions.

First, the features (FFT & Statistical) were extracted and exported, then the new data was downsampled and exported. In the next experiment, the original data was downsampled then features were extracted. With that, the datasets are as follow:

- Raw data.
- Statistical and FFT data without resampling.
- Data after extracting Statistical and FFT features then downsampled.
- Data after downsampling then extracting Statistical and FFT features.

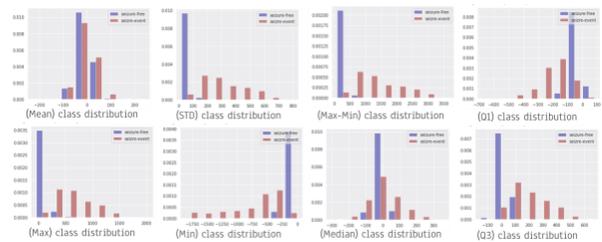


Figure 3. Class distribution for each of the extracted features from top left: (A) Mean, (B)Standard Deviation, (C) Max-Min, and (D) Q1, From down left: (E) Maximum value, (F) Minimum value, (G) Median, (H) Q3.

IV. EXPERIMENTAL DESIGN AND ANALYSIS

For this study, a dummy classifier and a baseline model (Bayesian Ridge Regression) was used on the original dataset, with an assumption that there is no data leakage, for instance all records are assumed to be independent and do not come from the same patient; therefore, it does not address it in advance. The overall performance of both of the classifiers are low.

Table 1. Performance of dummy classifier vs. 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

The next part focuses on time series classification with classical Machine learning models, it assumes that there is a possible source of data leakage and solves it

beforehand. In the 4th experiment BBR model was implemented on the fixed data after extracting the statistical and FFT features, which scored 75% recall and 99% AUC, and given that the data is imbalanced and AUC is not sensitive to imbalanced data, the overall performance is considered good. Next, the experiment explores the impact the order of feature extraction and resampling on the performance of the model. As illustrated below the performance didn't change with different orders. Comparing the two experiments the performance is close in terms of the measurements under study, which gives an insight about the nature of the data.

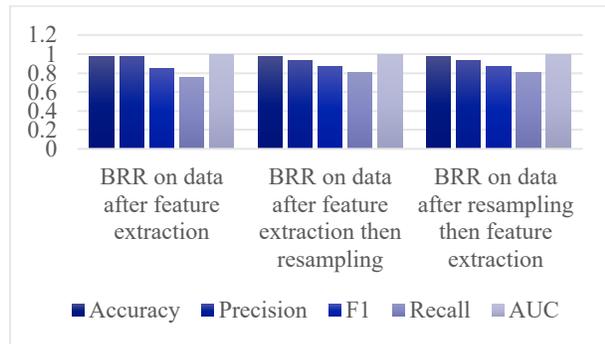


Figure 4. Performance of Approach 4,5 and 6.

The last part compares the performance of classical ML pipeline which involves feature engineering and an RNN model. As can be seen in Figure below, applying BBR on the data without feature engineering, underperformed the other approaches significantly, especially in terms of F1, Recall and AUC which are considered more important in this case. The performance of BBR with feature engineering is close to the performance of RNN. RNN recall outperformed the classical pipeline's recall, nonetheless, BBR's AUC with feature engineered set slightly outperformed RNN AUC.

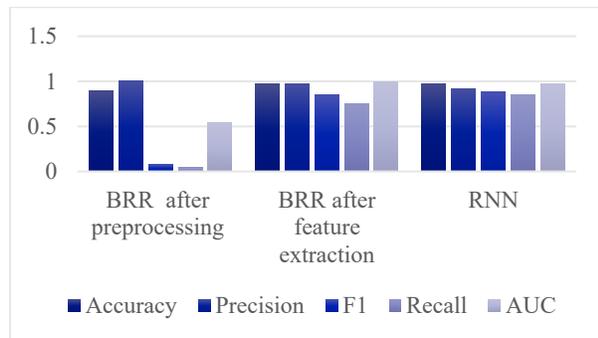


Figure 5. Performance of Approach 3,4 and 7.

V. Summary

In this paper, we have investigated the impact of resampling and its order in imbalanced time-series datasets. It also compared the performance of classical machine learning pipeline, which involved manual feature engineering and RNN.

Table 2. A comparison of performances

Approach	Recall	AUC	F1
Dummy Classifier	0.24	0.49	-
BBR on original data	0.02	0.5	0.04
BBR on data after addressing data leakage and class ratio in training and testing	0.04	0.54	0.08
BBR on data after feature extraction	0.75	0.996	0.85
BBR on data after feature extraction then resampling	0.81	0.995	0.87
BBR on data after resampling then feature extraction	0.81	0.995	0.87
RNN	0.85	0.97	0.88

The table shows the overall recall and AUC of the experiments. It can be seen that the recall of the dummy classifier performed better than the second and third experiments where only the data leakage and class ratio in training and testing was solved. The AUC in these three experiments were close. After introducing Feature engineering the recall improved dramatically in classical pipeline and slightly better in RNN where there is no (manual) feature engineering. On the other hand, the AUC value in the first 3 experiments didn't show a big difference comparing to each other, yet and again the performance increased after feature engineering in classical pipeline and RNN, exceeding 95%.

VI. CONCLUSION AND RECOMMENDATIONS

From above experiments, it was clear that the feature engineering improved the performance dramatically in the classical pipeline, and that the order of the feature engineering and resampling (downsampling) didn't affect the performance, and although applying a classical ML model on manually extracted features and RNN gave approximately close results, the recall which spots every seizure was better in RNN comparing to the classical approach.

Researchers working on similar problems are encouraged to explore different features derived from the dataset such as Lags, as well as different Deep learning architectures in classifying time-series data.

In this experiment, the class distribution was unbalanced in the training and testing sets, and further research is suggested to be conducted around the significance of unbalanced test in measuring the model's performance when the positive class is the minority one.

ACKNOWLEDGMENTS

I would like to express my gratitude to the Advanced Electronics Company AEC and Princess Nourah bint Abdulrahman University PNU. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of AEC or PNU.

REFERENCES

- [1] Epilepsy, World Health Organization. [Online]. Available: <https://www.who.int/news-room/factsheets/detail/epilepsy>. [Accessed: 24-Apr-2020].
- [2] Diagnosis. [Online]. Available: <https://www.nhs.uk/conditions/epilepsy/diagnosis/>. [Accessed: 24-Apr-2020].
- [3] Electroencephalogram (EEG). [Online]. Available: <https://www.nhs.uk/conditions/electroencephalogram/>. [Accessed: 24-Apr-2020].
- [4] Hussein, Ramy, Hamid, Ward, Wang, and Z. Jane, Epileptic Seizure Detection: A Deep Learning Approach, arXiv.org, 27-Mar-2018. [Online]. Available: <https://arxiv.org/abs/1803.09848>. [Accessed: 24-Apr-2020].
- [5] Zhou, et al. Epileptic Seizure Detection Based on EEG Signals and CNN. *Frontiers*, Frontiers, 23 Nov. 2018. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fninf.2018.00095/full>. [Accessed: 24-Apr-2020].
- [6] Roshan, et al. Epileptic EEG classification using nonlinear parameters on different frequency bands *Journal of Mechanics in Medicine and Biology*, February 2015. [Online]. Available: https://www.researchgate.net/publication/276394581_Epileptic_EEG_classification_using_nonlinear_parameters_on_different_frequency_bands. [Accessed: 24-Apr-2020].
- [7] Hamad, et al. A Hybrid EEG Signals Classification Approach Based on Grey Wolf Optimizer Enhanced SVMs for Epileptic Detection Springer International Publishing, Jan. 2018. [Online]. Available: https://www.researchgate.net/publication/319404363_A_Hybrid_EEG_Signals_Classification_Approach_Based_on_Grey_Wolf_Optimizer_Enhanced_SVMs_for_Epileptic_Detection. [Accessed: 24-Apr-2020].
- [8] Chen, Guangyi, et al. Automatic Epileptic Seizure Detection in EEG Using Nonsampled Wavelet Fourier Features. SpringerLink, Springer Berlin Heidelberg, 16 Jan. 2017. [Online]. Available: <https://link.springer.com/article/10.1007/s40846-016-0214-0>. [Accessed: 24-Apr-2020].
- [9] T. Fawcett, F. Provost, C. Phua, D. Alahakoon, V. Lee, M. Kubat, Holte, S. Matwin, G. Cui, Wong, Lui, C. Chen, A. Liaw, L. Breiman, Zhou, Liu, Batista, Prati, Y. Sun, M. S. Kamel, Wong, Y. Wang, Ting, Weiss, N. Japkovic, A. Kotcz, T. Hastie, R. Tibshirani, J. Friedman, J. Franklin, A. Asuncion, J. Newman, Ge, and Gm, Learning to improve medical decision making from imbalanced data without a priori cost, *BMC Medical Informatics and Decision Making*, 01-Jan-1997. [Online]. Available: <https://link.springer.com/article/10.1186/s12911-014-0111-9#Abs1>. [Accessed: 24-Apr-2020].
- [10] N. Maisarah M. Sobran, A. Ahmad, Zuwairie Ibrahim, CLASSIFICATION OF IMBALANCED DATASET USING CONVENTIONAL NAAVE BAYES CLASSIFIER, 26 November 2013, [Online]. Available: https://worldconferences.net/proceedings/aics2013/oc/papers_aics2013/A021%20-%20NUR%20MAISARAH%20MOHD%20SOBRAN%20-%20Classification%20of%20Imbalanced%20dataset%20using%20conventional%20naive%20bayes%20classifier.pdf [Accessed: 24-Apr-2020].
- [11] D.-C. Li, C.-W. Liu, and S. C. Hu, A learning method for the class imbalance problem with medical data sets, *Computers in Biology and Medicine*, 26-Mar-2010. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0010482510000405>. [Accessed: 24-Apr-2020].
- [12] R. Batuwita and V. Palade, "Efficient resampling methods for training support vector machines with imbalanced datasets," *IJCNN*, 2010. [Online]. Available: <https://ieeexplore.ieee.org/document/5596787>. [Accessed: 24-Apr-2020].
- [13] D.-C. Li, C.-W. Liu, and S. C. Hu, A learning method for the class imbalance problem with medical data sets, *Computers in Biology and Medicine*, 26-Mar-2010. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0010482510000405>. [Accessed: 24-Apr-2020].
- [14] S. Amari, S. Wu, C. Cortes, V. Vapnik, J. Demšar, Y. Freund, H. Guo, Viktor, J. Laurikkala, N. Japkowicz, S. Stephen, M. Kukar, C. Nadeau, Y. Bengio, S. Rosset, C. Perlich, B. Zadrozny, Ting, V. Vapnik, H. Zhao, Zhou, and Liu, Boosting support vector machines for imbalanced data sets, *Knowledge and Information Systems*, 01-Jan-1999. [Online]. Available: <https://link.springer.com/article/10.1007/s10115-009-0198-y>. [Accessed: 24-Apr-2020].
- [15] J. Li, B. Zhang, B. Zhang, U. Wien, Jinhui Yuan Tsinghua University, Jinhui Yuan, Tsinghua University, Jianmin Li Tsinghua University, Tsinghua University, Uuc, Ucsb, Microsoft Research, and Unc, Learning concepts from large scale imbalanced data sets using support cluster machines, Learning concepts from large scale imbalanced data sets using support cluster machines | *Proceedings of the 14th ACM international conference on Multimedia*, 01-Oct-2006. [Online]. Available: <https://dl.acm.org/doi/10.1145/1180639.1180729>. [Accessed: 24-Apr-2020].

- [16] S. Maldonado, R. Weber, and F. Famili, "Feature selection for high-dimensional class-imbalanced data sets using Support Vector Machines," *Information Sciences*, 30-Jul-2014. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0020025514007154?via=ihub>. [Accessed: 24-Apr-2020].
- [17] UCI Machine Learning Repository: Epileptic Seizure Recognition Data Set. [Online]. Available: [https://archive.ics.uci.edu/ml/datasets/Epileptic Seizure Recognition](https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition). [Accessed: 24-Apr-2020].
- [18] S. Kaufman, S. Rosset, C. Perlich, "Leakage in Data Mining: Formulation, Detection, and Avoidance," *Shachar ACM Trans. Knowl. Discov. Data*, 2012. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.365.7769&rep=rep1&type=pdf>. [Accessed: 24-Apr-2020].
- [19] Dansbecker, "Data Leakage," *Kaggle*, 22-Jan-2018. [Online]. Available: <https://www.kaggle.com/dansbecker/data-leakage>. [Accessed: 24-Apr-2020].
- [20] scipy.stats.kruskal, *scipy.stats.kruskal - SciPy v1.4.1 Reference Guide*. [Online]. Available: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kruskal.html>. [Accessed: 24-Apr-2020].
- [21] "Fast Fourier Transform, Fast Fourier Transform - an overview | ScienceDirect Topics." [Online]. Available: <https://www.sciencedirect.com/topics/engineering/fast-fourier-transform>. [Accessed: 24-Apr-2020].