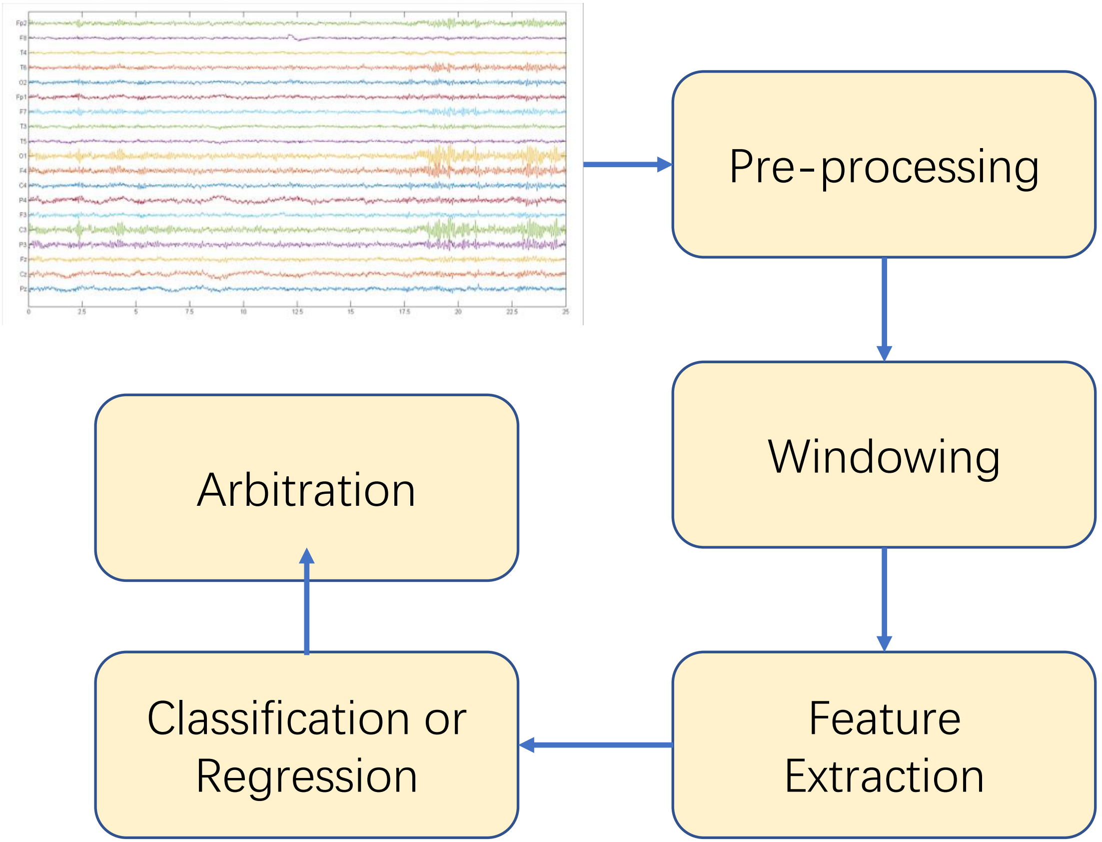
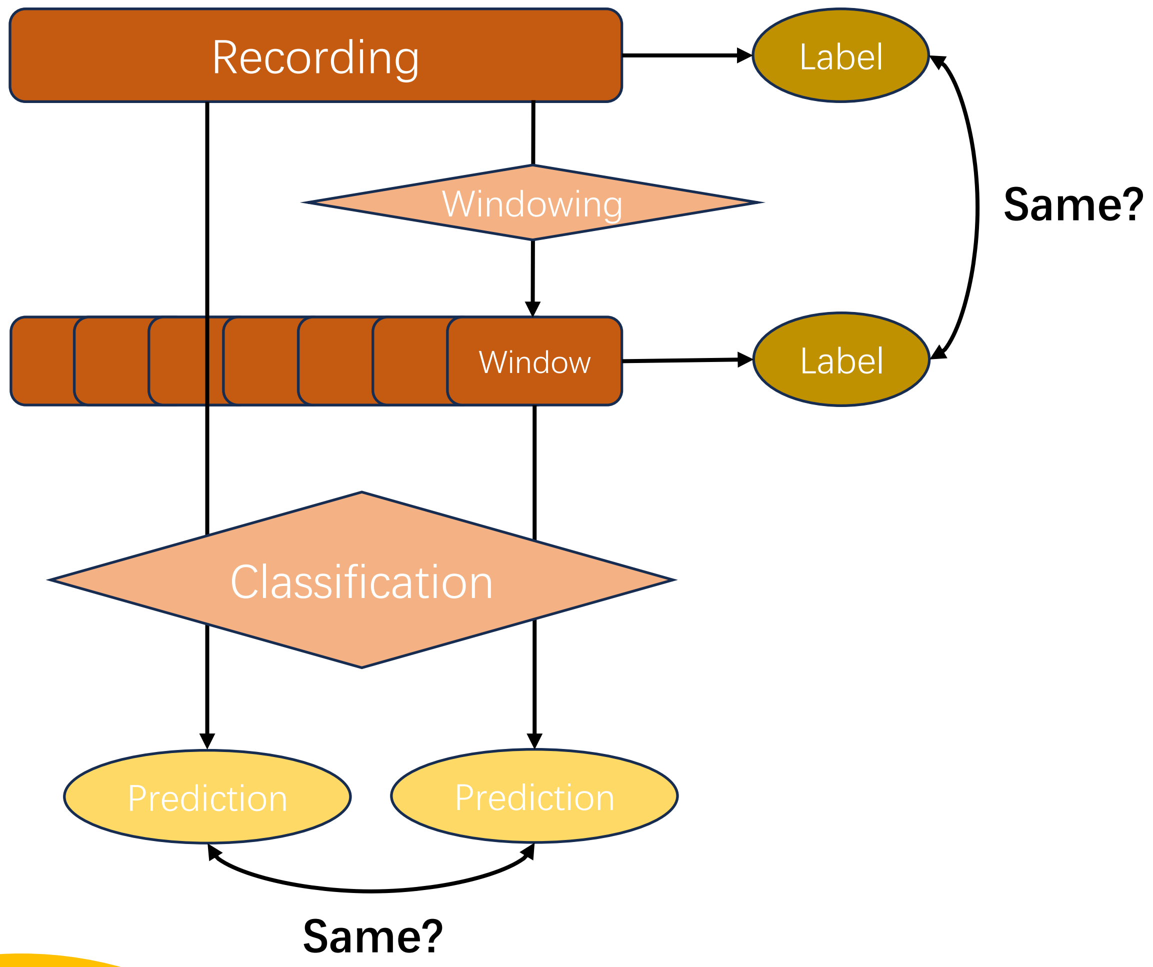
Scope and Arbitration in Machine Learning Clinical EEG Classification

Yixuan Zhu yixuan2.zhu@live.uwe.ac.uk
Luke J. W. Canham luke.canham@nbt.nhs.uk
David Western david.western@uwe.ac.uk

Introduction

A typical EEG classification task workflow based on machine learning.

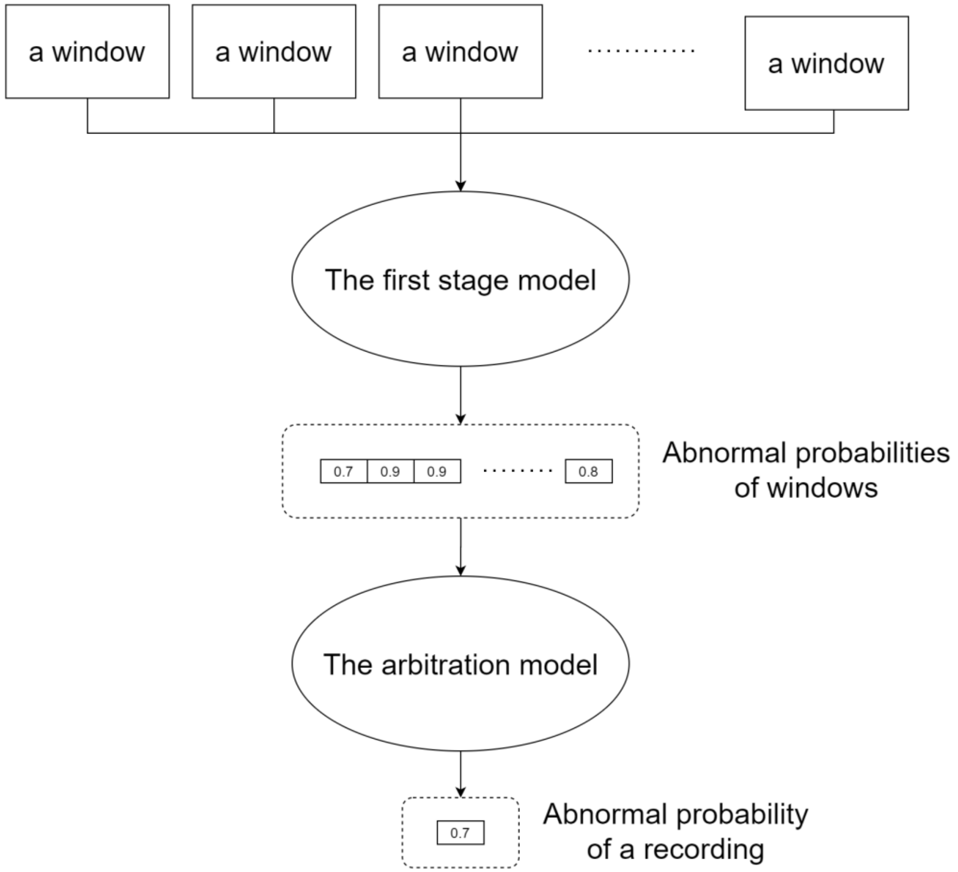




The Windowing Conundrum in EEG Classification

- **Labeling Dilemma:** One label per session may not reflect the true nature of each window, leading to potential misclassification.
- **Result Aggregation:** Determining the overall recording status from windowed results is uncertain—abnormal recordings could contain normal windows.

Method

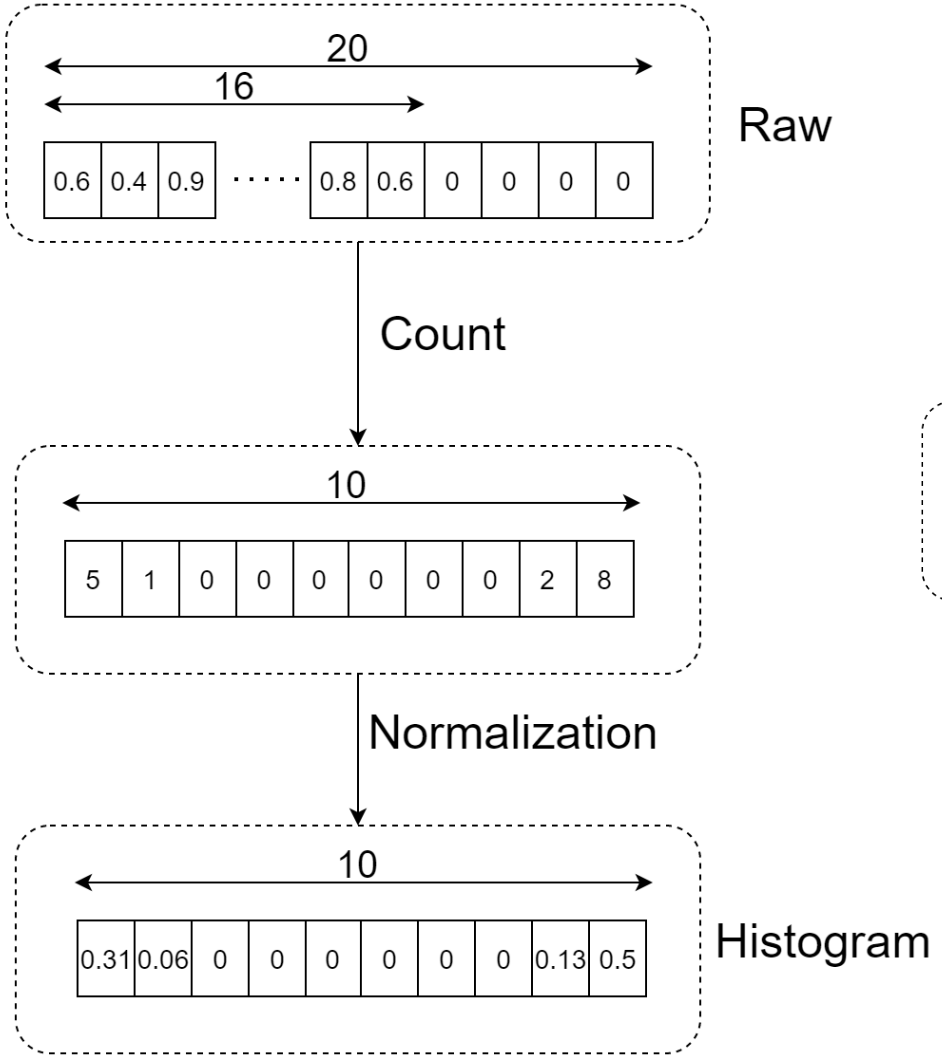


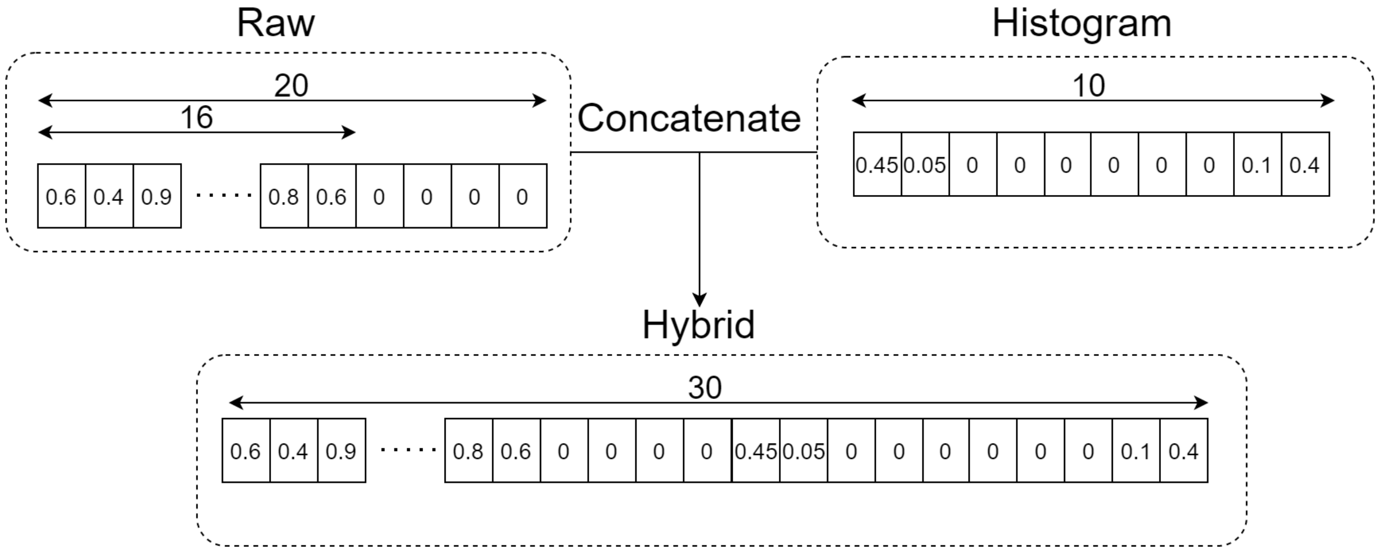
Integrating EEG Data with an Arbitration Model

Model Design: We created a model to synthesize window results in an EEG.
Window-Level Learning: It captures window interrelationships.
Bias Correction: It corrects label bias for accurate EEG interpretation.
Arbitration Innovation: First to emphasize and utilize machine learning in arbitration, marking a significant advancement in the field.

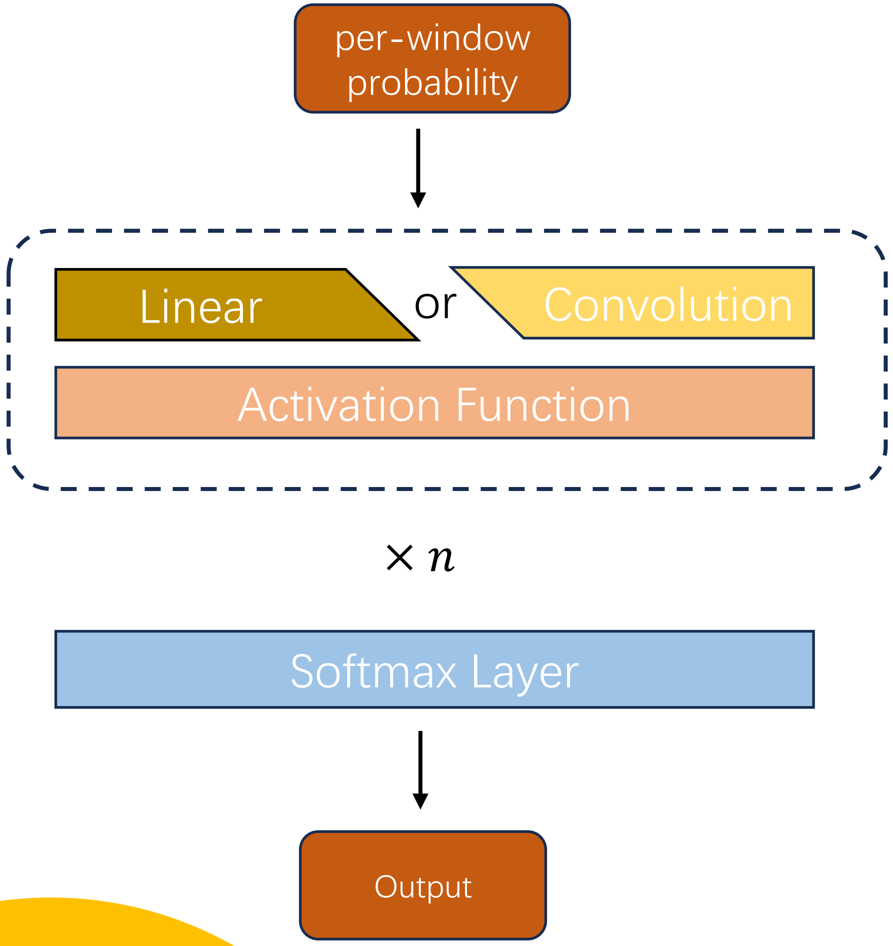
UWE Bristol University of the West of England

Diverse Input Strategies for EEG Data Analysis





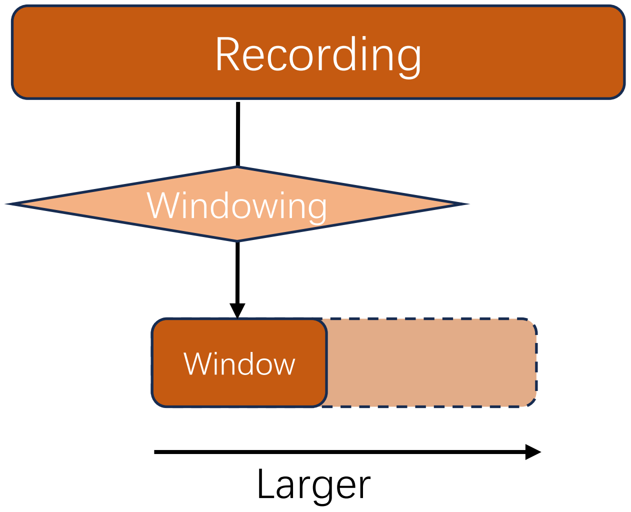
UWE Bristol University of the West of England



Arbitration Model Architecture (Artificial Neural Network)

- **Final Architecture:** The proposed models consist of a fully-connected layer followed by a softmax layer for final classification.
- **Depth and Size Variations:** We tested multi-layer perceptrons with 1 to 4 layers and hidden layers ranging from 5 to 20 units in size.
- **Alternatives to Fully-Connected Layers:** We also experimented with convolutional layers as an alternative to fully-connected ones.
- **Activation Functions:** Different activation functions were trialed, including RELU, ELU, and GELU.

The Scope Issue and Window Length



Scope Limitation: Traditional models may miss the full EEG context.
Global Info: Bigger windows provide a broader view.
Solution: Longer windows for better EEG capture.
Note: This complements machine learning arbitration.

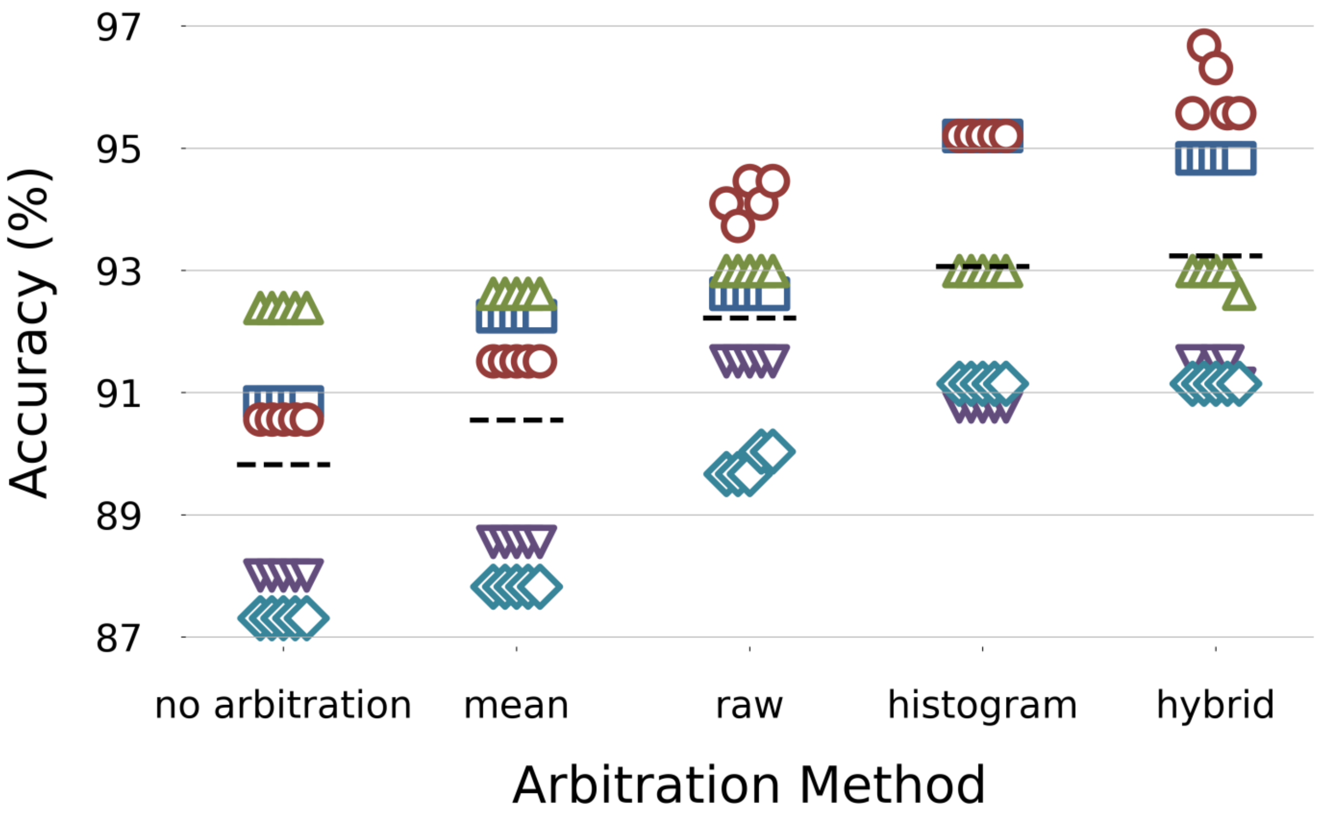
Results

Results

Table 1. Summary of state-of-the-art performance metrics for different models applied to abnormal EEG classification

Model	Accuracy	Sensitivity	Specificity
1D-CNN (T5-O1 channel)[12]	79.3 %	71.4 %	86.0 %
1D-CNN (F4-C4 channel)[12]	74.4 %	55.6 %	90.7 %
Deep4 [1]	85.4 %	75.1 %	94.1 %
TCN [5]	86.2 %		
ChronoNet [9]	86.6 %		
Alexnet[2]	87.3 %	78.6 %	94.7 %
VGG-16 [2]	86.6 %	77.8 %	94.0 %
Fusion Alexnet[8]	89.1 %	80.2 %	96.7 %
[6]	89.8 %	81.3 %	**96.9 %**
Proposed	**93.3 %**	**92.0 %**	92.9 %

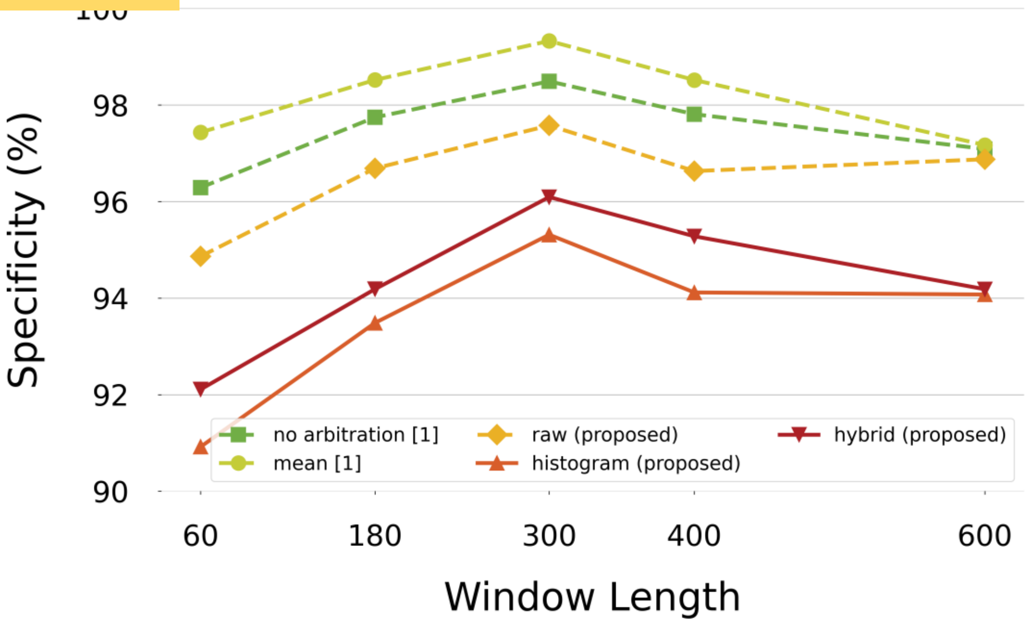
Arbitration Model Results on BD-Deep4 [1]

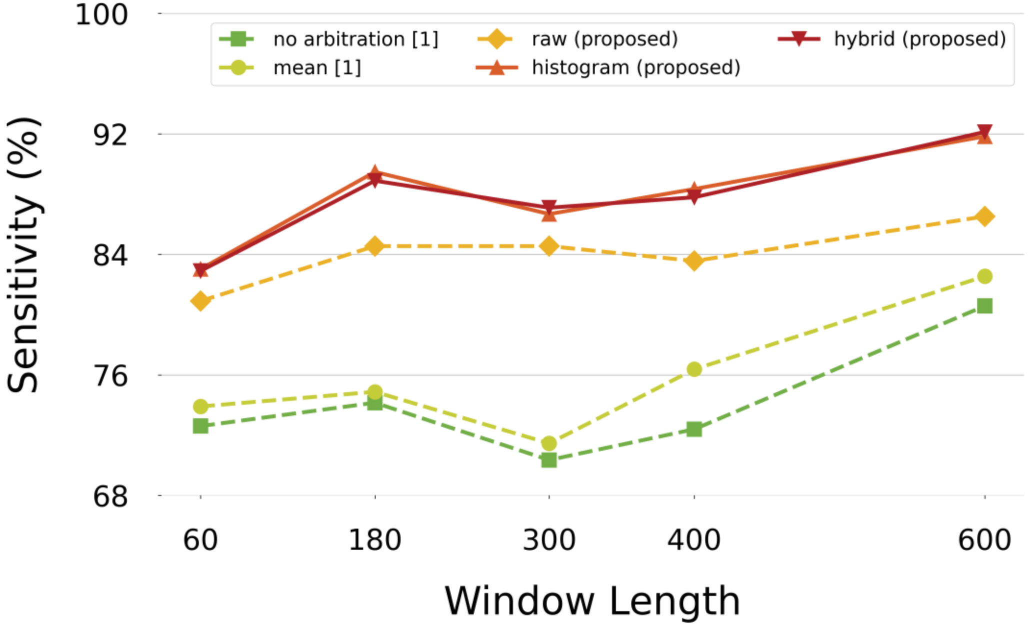


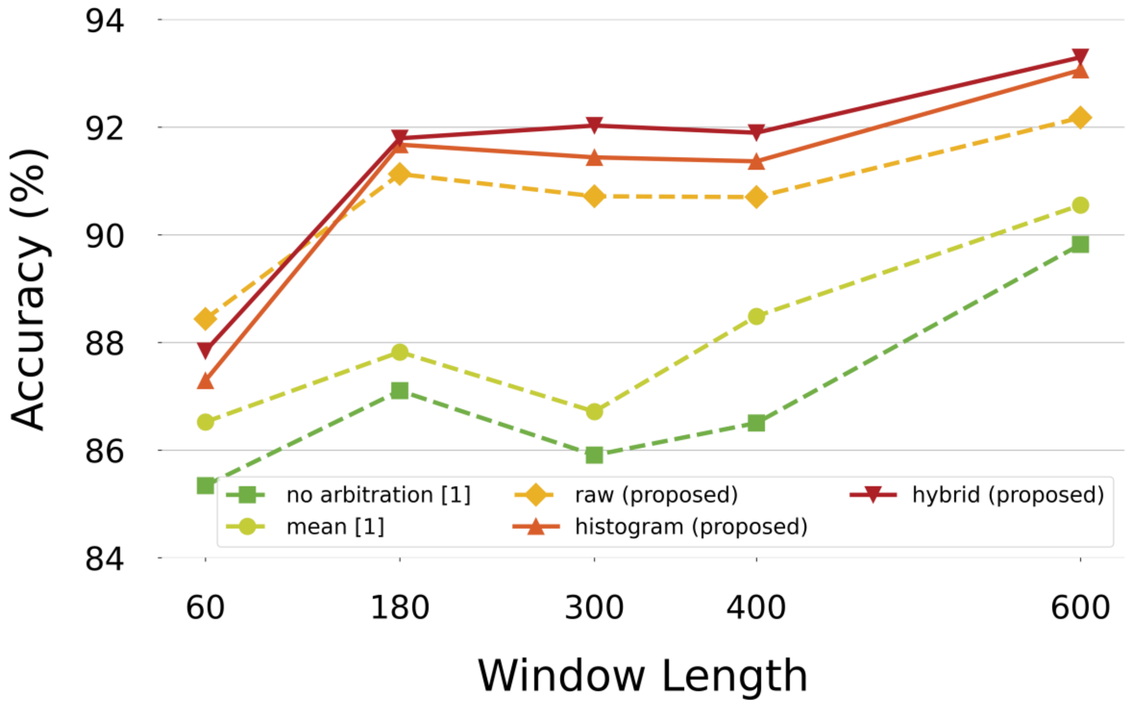
Performance of different arbitration models using window lengths of 60 s. Points with the same marker shape come from the same instance of the first-stage model (BD-Deep4). The dashed lines represent the mean for each arbitration method.

Effect of Window Length

Increasing window length improves accuracy by increasing sensitivity, with relatively little effect on specificity

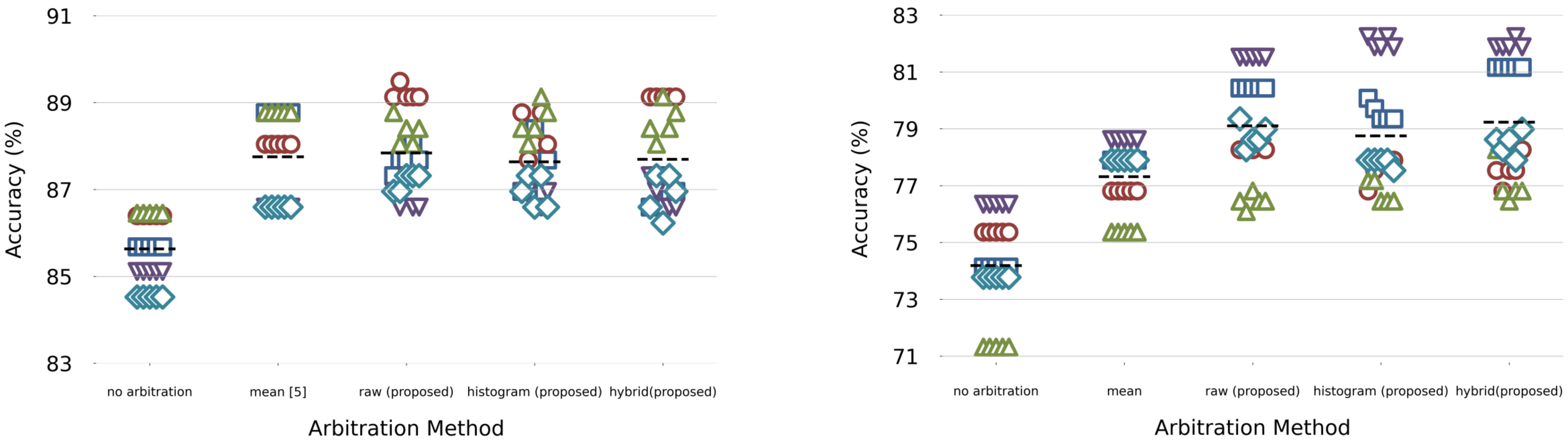






UWE Bristol University of the West of England

Arbitration Model Results on TCN [5,14] and ViT [15]



(a) TCN (b) ViT

Performance of different arbitration methods using (a) TCN and (b) ViT as the first-stage architecture with a window length of 60 s.

Conclusion

Conclusion

- **Approach Benefits:** Our method surpasses prior EEG classification benchmarks.
- **Enhanced Sensitivity:** More accurate window-label alignment boosts model sensitivity.
- **Clinical Relevance:** This could streamline EEG analysis for healthcare providers.
- **Broader Impact:** The method may be applicable to other time-series tasks.
- **Note:** The inter-rater agreement ceiling is not a constraint for TUAB.

Thank You!

Yixuan Zhu yixuan2.zhu@live.uwe.ac.uk
Luke J. W. Canham luke.canham@nbt.nhs.uk
David Western david.western@uwe.ac.uk

References

References

[1] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, “Deep learning with convolutional neural networks for eeg decoding and visualization,” Human brain mapping, vol. 38, no. 11, pp. 5391–5420, 2017.
[2] S. U. Amin, M. S. Hossain, G. Muhammad, M. Alhussein, and M. A. Rahman, “Cognitive smart healthcare for pathology detection and monitoring,” IEEE Access, vol. 7, pp. 10 745– 10 753, 2019.
[3] H. Banville, O. Chehab, A. Hyvärinen, D.-A. Engemann, and A. Gramfort, “Uncovering the structure of clinical eeg signals with self-supervised learning,” Journal of Neural Engineering, vol. 18, no. 4, p. 046020, 2021.
[4] H. Banville, S. U. Wood, C. Aimone, D.-A. Engemann, and A. Gramfort, “Robust learning from corrupted eeg with dynamic spatial filtering,” NeuroImage, vol. 251, p. 118994, 2022.
[5] L. A. Gemein, R. T. Schirrmeister, P. Chrab ˛aszcz, D. Wilson, J. Boedecker, A. Schulze-Bonhage, F. Hutter, and T. Ball, “Machine-learning-based diagnostics of eeg pathology,” NeuroImage, vol. 220, p. 117021, 2020.
[6] G. Muhammad, M. S. Hossain, and N. Kumar, “Eeg-based pathology detection for home health monitoring,” IEEE Journal on Selected Areas in Communications, vol. 39, no. 2, pp. 603– 610, 2020.
[7] N. Wagh and Y. Varatharajah, “Eeg-gcnn: Augmenting electroencephalogram-based neurological disease diagnosis using a domain-guided graph convolutional neural network,” Machine Learning for Health. PMLR, 2020, pp. 367–378.
[8] M. Alhussein, G. Muhammad, and M. S. Hossain, “Eeg pathology detection based on deep learning,” IEEE Access, vol. 7, pp. 27 781–27 788, 2019

UWE Bristol University of the West of England

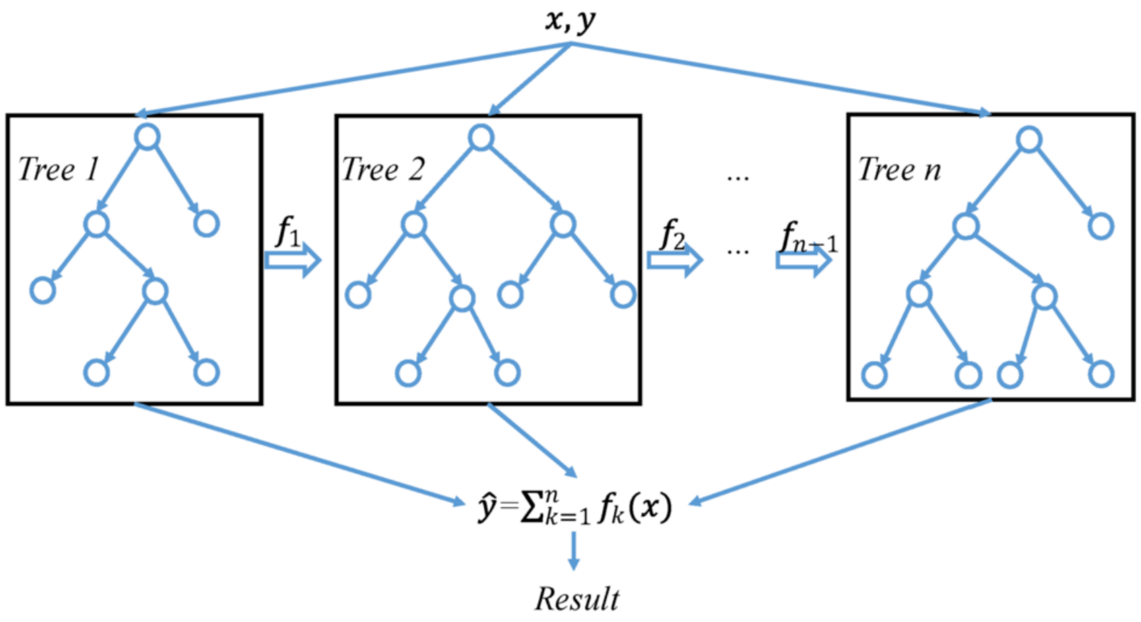
References

[9] S. Roy, I. Kiral-Kornek, and S. Harrer, “Chrononet: a deep recurrent neural network for abnormal eeg identification,” Artificial Intelligence in Medicine: 17th Conference on Artificial Intelligence in Medicine, AIME 2019, Poznan, Poland, June 26– 29, 2019, Proceedings 17. Springer, 2019, pp. 47–56.
[10] S. López, I. Obeid, and J. Picone, “Automated interpretation of abnormal adult electroencephalograms,” Ph.D. dissertation, 2017.
[11] I. Obeid and J. Picone, “The temple university hospital eeg data corpus,” Frontiers in neuroscience, vol. 10, p. 196, 2016.
[12] Ö. Yıldırım, U. B. Baloglu, and U. R. Acharya, “A deep convolutional neural network model for automated identification of abnormal eeg signals,” Neural Computing and Applications, vol. 32, pp. 15 857–15 868, 2020.
[13] D. Western, T. Weber, R. Kandasamy, F. May, S. Taylor, Y. Zhu, and L. Canham, “Automatic report-based labelling of clinical eegs for classifier training,” 2021 IEEE Signal Processing in Medicine and Biology Symposium (SPMB). IEEE, 2021, pp. 1– 6.
[14] S. Bai, J. Z. Kolter, and V. Koltun, “An empirical evaluation of generic convolutional and recurrent networks for sequence modeling,” 2018. (available at: https://arxiv.org/abs/1803. 01271).
[15] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” arXiv preprint arXiv:2010.11929, 2020.
[16] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier neural networks,” Proceedings of the fourteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings, 2011, pp. 315–323.

Appendix

Arbitration Model Architecture (XGBoost)

- **Tabular Data Perspective:** We treated the outputs of the first-stage model as tabular data, which led us to evaluate XGBoost, known for its strong performance on such datasets.
- **Hyperparameter Optimization:** A grid search was utilized during each training run to find the optimal 'maximum depth' from {5, 10, 15, 20, 25}, using three-fold cross-validation for selection.



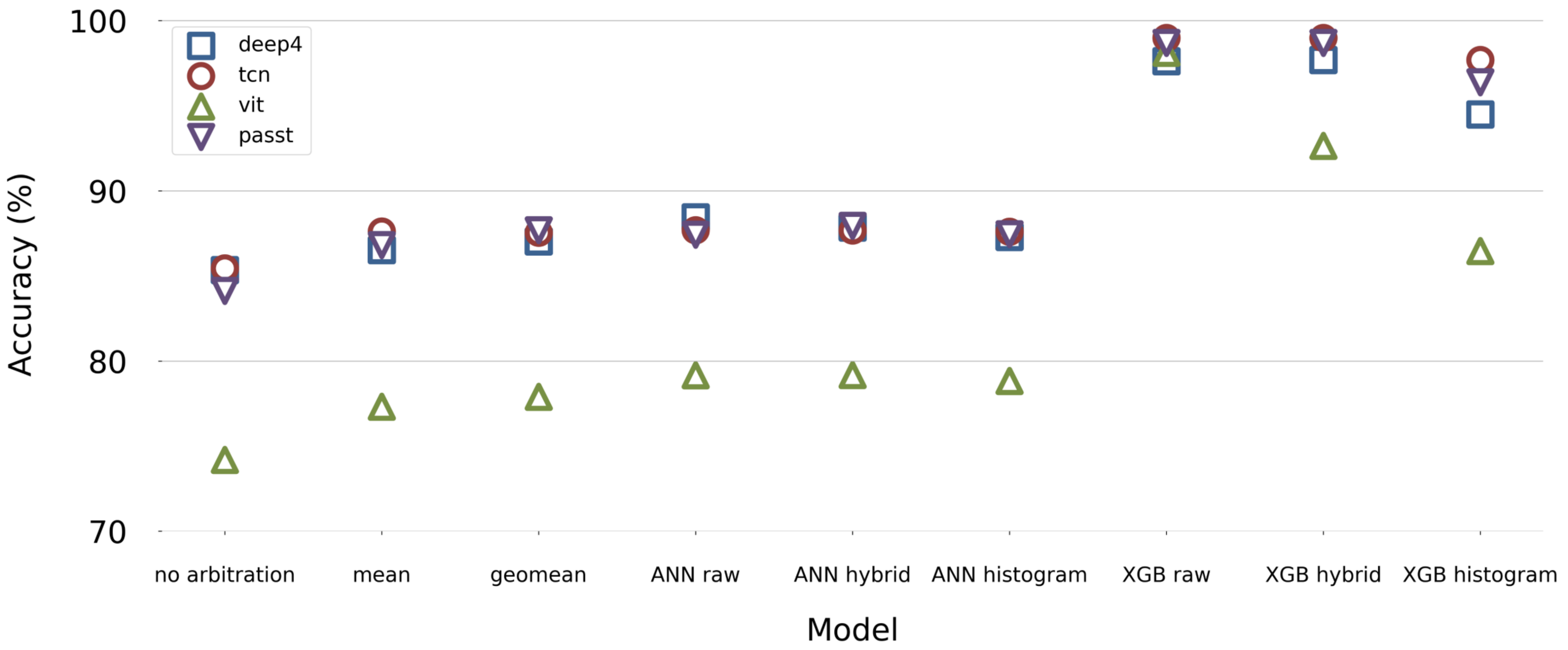
Comparison of performance between models

Model	Accuracy	Sensitivity	Specificity
1D-CNN (T5-O1 channel) (Yıldırım et al., 2020)	79.3 %	71.4 %	86.0 %
1D-CNN (F4-C4 channel) (Yıldırım et al., 2020)	74.4 %	55.6 %	90.7 %
Deep4 (Schirrmeister et al., 2017)	85.4 %	75.1 %	94.1 %
TCN (Gemein et al., 2020)	86.2 %		
ChronoNet (Roy et al., 2019)	86.6 %		
Alexnet (Amin et al., 2019)	87.3 %	78.6 %	94.7 %
VGG-16 (Amin et al., 2019)	86.6 %	77.8 %	94.0 %
Fusion Alexnet (Alhussein et al., 2019)	89.1 %	80.2 %	96.7 %
Fusion CNN (Muhammad et al., 2020)	89.8 %	81.3 %	96.9 %
Scope and Arbitration (Deep4-ANN-Hybrid) Zhu et al. (2023)	93.3 %	92.0 %	92.9 %
Window Stacking Meta-Model (TCN-XGBoost-Raw)	**99.0 %**	**98.1 %**	**100 %**

Table 1: Summary of state-of-the-art performance metrics for different models applied to abnormal EEG classication on the TUAB dataset

UWE Bristol University of the West of England

Second-Stage Model Results



Performance comparison of single-stage and various two-stage architectures, all with a window length of 60s, using the TUAB dataset. Each column represents a different arbitration method. Each marker type represents a different first-stage architecture. Each data point is the average accuracy across twenty-five experiments.

UWE Bristol University of the West of England